

**SVEUČILIŠTE U ZAGREBU**  
**PRIRODOSLOVNO–MATEMATIČKI FAKULTET**  
**MATEMATIČKI ODSJEK**

Ante Šterc

**SVOJSTVENE VRIJEDNOSTI**  
**KOVARIJACIJSKIH MATRICA ZA**  
**RAZDIOBE TEŠKIH REPOVA**

Diplomski rad

Voditelj rada:  
prof. dr. sc. Bojan Basrak

Zagreb, svibanj, 2018.

Ovaj diplomski rad obranjen je dana \_\_\_\_\_ pred ispitnim povjerenstvom u sastavu:

1. \_\_\_\_\_, predsjednik
2. \_\_\_\_\_, član
3. \_\_\_\_\_, član

Povjerenstvo je rad ocijenilo ocjenom \_\_\_\_\_.

Potpisi članova povjerenstva:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

*Mami i tati.*

# Sadržaj

<b>Sadržaj</b>	<b>iv</b>
<b>Uvod</b>	<b>2</b>
<b>1 Pojmovi, definicije i poznati rezultati</b>	<b>3</b>
1.1 Slučajne matrice . . . . .	3
1.2 Uzoračke kovarijacijske matrice . . . . .	4
1.3 Granični rezultati za ekstremne svojstvene vrijednosti . . . . .	7
1.4 Svojstveni vektori . . . . .	10
1.5 Teški repovi . . . . .	13
1.6 Konvergencija po distribuciji slučajnih mjera . . . . .	13
<b>2 Svojstvene vrijednosti kovarijacijske matrice</b>	<b>16</b>
2.1 Temeljne aproksimacije . . . . .	17
2.2 Svojstveni vektori . . . . .	20
2.3 Autokovarijacijske matrice . . . . .	23
<b>3 Uzoračke autokovarijacijske matrice</b>	<b>26</b>
3.1 Konvergencija točkovnih procesa . . . . .	29
3.2 Zavisnost između redaka i stupaca . . . . .	33
<b>4 Primjene</b>	<b>37</b>
4.1 Uzoračke kovarijacijske matrice . . . . .	37
4.2 Analiza podataka indeksa S&P 500 . . . . .	43
4.3 Sume kvadrata uzoračke autokovarijacijske matrice . . . . .	47
<b>Bibliografija</b>	<b>51</b>

# Uvod

U novije vrijeme, razvojem tehnologije, skupovi podataka koji se obrađuju su sve veći i kompleksniji. U mnogim granama znanosti kao što su na primjer meteorologija, genetika, telekomunikacije, kvantna fizika..., kao i pri obradi podataka o društvenim mrežama, financijskim podacima itd., javlja se potreba za obradom velike količine podataka da bi se izvukle bitne i korisne informacije u vrlo kratkom vremenu. Matrice podataka velikih dimenzija najčešće su komplicirane strukture i prevelike da bi se obradile klasičnim metodama. Stoga se za proučavanje strukture matrica često koriste uzoračke kovarijacijske i korelacijske matrice. Često korištena tehnika za smanjivanje dimenzije podataka je metoda glavnih komponenti (eng. *principal component analysis* (PCA)) koja promatra najveće svojstvene vrijednosti i pripadne svojstvene vektore uzoračke kovarijacijske matrice.

Teorija o slučajnim matricama (eng. *Random Matrix Theory* (RMT)) bavi se svojstvima slučajnih matrica velikih dimenzija. Povijesno, a i danas, razvoj teorije je potaknut zahtjevima iz primjena. Primjerice, razina energije čestica u velikom sustavu može se opisati svojstvenim vrijednostima beskonačnodimenzionalnih slučajnih Hermiteovih operatora  $W$  u Hilbertovom prostoru. Zbog ograničenja, problem se vrlo često diskretizira u konačnodimenzionalnom prostoru. U tom slučaju  $W$  postaje Hermiteova slučajna matrica velike, ali konačne dimenzije. Takve matrice zovu se *Wignerove* matrice po poznatom matematičaru E. P. Wigneru. On je dokazao sljedeću tvrdnju: ako je distribucija komponenti matrice  $W$  na dijagonali i iznad glavne dijagonale standardna normalna, tada očekivana empirijska distribucija svojstvenih vrijednosti matrice  $W$  teži zakonu polukruga, kako dimenzija matrice  $W$  raste u beskonačnost.

Asimptotsko ponašanje najvećih svojstvenih vrijednosti slučajnih kovarijacijskih matrica rastućih dimenzija u početku je proučavano u kontekstu tzv. gaussovskih matrica. Johnston je u radu [16] 2001. godine dokazao teorem o ponašanju najveće svojstvene vrijednosti. Taj rezultat su nešto kasnije, 2010. godine, proširili Tao i Vu koristeći metodu koja se bazira na tzv. teoremu četiri momenta (eng. *Four Moment Theorem*) u radu [24]. U slučaju razdioba s beskonačnim četvrtim momentom, razvoj je išao nešto sporije te je inicijalni rezultat pokazao Soshnikov u radu [22] 2004. godine. Razvoju u tom smjeru posljednjih godina izniman doprinos dali su među ostalima Davis, Heiny i Mikosch.

Ovaj rad započinje definicijama i pregledom nekih dosadašnjih rezultata iz ovog po-

dručja u slučaju lakih, kao i u slučaju razdioba teških repova. U drugom poglavlju prikazana su neka od najnovijih istraživanja Davisa, Heinyja, Mikoscha i njihovih suradnika o asimptotskim svojstvima najvećih svojstvenih vrijednosti uzoračkih kovarijacijskih matrica vremenskih nizova s distribucijom teških repova i nezavisno jednako distribuiranim komponentama.

U trećem poglavlju prikazano je kako se asimptotska konvergencija najveće svojstvene vrijednosti uzoračke kovarijacijske matrice može dobiti iz konvergencije odgovarajućeg točkovnog procesa kao i neki rezultati u slučaju kada postoji zavisnost između redaka i stupaca matrice podataka. U posljednjem, četvrtom poglavlju ilustrirani su teorijski rezultati iz prethodnih poglavlja na simuliranim i empirijskim podacima te je provedena kratka analiza vremenskih nizova komponenti poznatog S&P 500 dioničkog indeksa.

# Poglavlje 1

## Pojmovi, definicije i poznati rezultati

### 1.1 Slučajne matrice

#### Granična spektralna distribucija

**Definicija 1.1.1.** Za svaku slučajnu  $p \times p$  matricu  $A$  sa realnim svojstvenim vrijednostima  $\lambda_1(A), \lambda_2(A), \dots, \lambda_p(A)$  uzoračka spektralna distribucija definirana je relacijom

$$F_A(x) = \frac{1}{p} \sum_{n=0}^p \mathbf{I}_{\{\lambda_i(A) \leq x\}}, \quad x \in \mathbb{R}.$$

Centralni problem u teoriji slučajnih matrica je odrediti tzv. *graničnu spektralnu distribuciju* niza  $(F_{A_n})$  za zadani niz Hermiteovih  $p \times p$  matrica  $(A_n)$ . Pod time se misli na konvergenciju po distribuciji niza vjerojatnosnih distribucija  $(F_{A_n})$  prema vjerojatnosnoj distribuciji  $F$ , tj.  $\lim_{n \rightarrow \infty} F_{A_n}(x) = F(x)$  g.s. za svaku točku neprekidnosti od  $F$ . Kako ne postoji formula u zatvorenom obliku ako je dimenzija matrice veća od 4 razvijaju su se metode koje karakteriziraju i identificiraju graničnu spektralnu distribuciju. Dvije najčešće korištene metode su metoda momenata i Stieltjesova transformacija koje i navodimo u nastavku.

Prvom metodom po lemi B.3 u [3] funkcija distribucije  $F$  ima jedinstvenu karakterizaciju preko niza svojih momenata

$$\beta_k = \int_0^\infty x^k dF(x), \quad k = 1, 2, \dots$$

ako je zadovoljen tzv. Carlemanov uvjet  $\sum_{k=1}^\infty \beta_{2k}^{-1/(2k)} = \infty$  (vidi npr. [1]). U ovom slučaju konvergencija po distribuciji niza  $(F_{A_n})$  prema  $F$  ekvivalentna je konvergenciji momenata

$$\beta_k(A_n) = \int_0^\infty x^k dF_{A_n}(x) = \frac{1}{p_n} \text{tr}(A_n^k) \rightarrow \beta_k, \quad n \rightarrow \infty, k = 1, 2, \dots \quad (1.1)$$

Često je računanje  $\text{tr}(\mathbf{A}_n^k)$  zahtjevno pa računanje očekivanja i varijance zahtijeva kombinatorne tehnike. S pozitivne strane, ako  $F$  ima konačan nosač, Carlemanov uvjet je automatski ispunjen.

Drugi pristup je korištenje Stieltjesove transformacije empirijske spektralne distribucije  $F_{\mathbf{A}}$

$$s_{F_{\mathbf{A}}}(z) = \int_{\mathbb{R}} \frac{1}{x-z} dF_{\mathbf{A}}(x) = \frac{1}{p} \text{tr}((\mathbf{A} - z\mathbf{I})^{-1}), \quad z \in \mathbb{C}^+,$$

gdje je  $\mathbb{C}^+$  oznaka za sve kompleksne brojeve s pozitivnim imaginarnim dijelom. Korištenjem Stieltjesove transformacije, konvergencija po distribuciji niza  $(F_{\mathbf{A}_n})$  prema  $F$  ekvivalentna je sa  $s_{F_{\mathbf{A}}}(z) \rightarrow s_F(z)$  g.s. za svaki  $z \in \mathbb{C}^+$ . Također, funkcija distribucije  $F$  određena je Stieltjesovom transformacijom  $s_F$  u svim točkama neprekidnosti  $a, b$  od  $F$  graničnom relacijom

$$F(b) - F(a) = \lim_{v \rightarrow 0^+} \frac{1}{\pi} \int_a^b \Im s_F(x + iv) dx,$$

gdje je  $\Im$  oznaka za imaginarni dio broja  $z \in \mathbb{C}$ .

## 1.2 Uzoračke kovarijacijske matrice

**Definicija 1.2.1.** Za  $n$ -dimenzionalni uzorak  $p$ -dimenzionalnog vremenskog niza koji je dan vektorima  $x_1, x_2, \dots, x_n$  uzoračka kovarijacijska matrica (nenormalizirana) definirana je kao

$$\mathbf{S} = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t = \sum_{i=1}^n x_i x_i^t - n\bar{x}\bar{x}^t = \mathbf{X}\mathbf{X}^t - n\bar{x}\bar{x}^t,$$

gdje je  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$  aritmetička sredina, a

$$\mathbf{X} = (x_1, x_2, \dots, x_n) = (X_{it})_{i=1,2,3,\dots,p; t=1,2,3,\dots,n}$$

je matrica podataka.

Uzoračka kovarijacijska matrica igra bitnu ulogu u višedimenzionalnoj statistici jer je od krucijalne važnosti u analizama kao što su: analiza glavnih komponenti (eng. *principal component analysis*), višedimenzionalna regresija, faktorska analiza, diskriminantna analiza i testiranje hipoteza. Važnu ulogu u razvijanju statističke teorije su imala višedimenzionalna normalno distribuirana opažanja. Uzoračke kovarijacijske matrice s normalno distribuiranim podacima bile su predmet Wishartovog proučavanja [28] još 1928. godine. U dvadesetom stoljeću procedure koje nisu asimptotske su razvijene za gaussovske obzervacije kao što su Fisherov  $F$ -test, Studentov  $t$ -test i analiza varijance. U praksi su opažanja rijetko normalno distribuirana pa se koriste asimptotske metode koje su bazirane na graničnim teoremima umjesto egzaktnih rezultata do kojih je teško doći.



Klasični granični teoremi koriste pretpostavku o fiksnoj dimenziji  $p$  dok dimenzija uzorka  $n$  raste u beskonačnost. Ako su pretpostavke *Jakog zakona velikih brojeva* ispunjene,  $n^{-1}\mathbf{S}$  konvergira g.s. prema kovarijacijskoj matrici  $\Sigma$  od  $x_1$ .

Kod klasičnih metoda, za velike  $p$ ,  $n^{-1}\mathbf{S}$  nije dobar procjenitelj za  $\Sigma$  i potrebne su prilagodbe. U suprotnom, korištenjem klasičnih metoda i za fiksni  $p$  za obradu podataka velikih dimenzija mogle bi voditi na krive zaključke.

Teorija slučajnih matrica pruža granične rezultate u slučaju velike dimenzije  $p$  na kojima se mogu graditi nove statističke metode. U teoriji slučajnih matrica pretpostavljamo da  $p = p_n$  raste s  $n$ . Uvjet koji se najčešće koristi u literaturi je

$$\lim_{n \rightarrow \infty} \frac{p_n}{n} = \gamma \in (0, \infty). \quad (1.2)$$

Asimptotsko spektralno ponašanje dvije velike matrice smatrat ćemo istim ako je njihova razlika konačnog ranga. U tu svrhu, za *uzoračku kovarijacijsku matricu*, od sada ćemo koristiti oznaku  $\mathbf{X}\mathbf{X}^t$ . Po teoremu A.44 u [3] za supremum normu vrijedi

$$\|\mathbf{F}_{n^{-1}\mathbf{S}} - \mathbf{F}_{n^{-1}\mathbf{X}\mathbf{X}^t}\| \leq p^{-1} \text{rang}((\bar{x}, \bar{x}, \dots, \bar{x})),$$

što znači da se granične spektralne distribucije podudaraju pod uvjetom (1.2), pa b.s.o.m.p. da su elementi matrice  $(X_{it})$  centrirani pod uvjetom da očekivanje postoji.

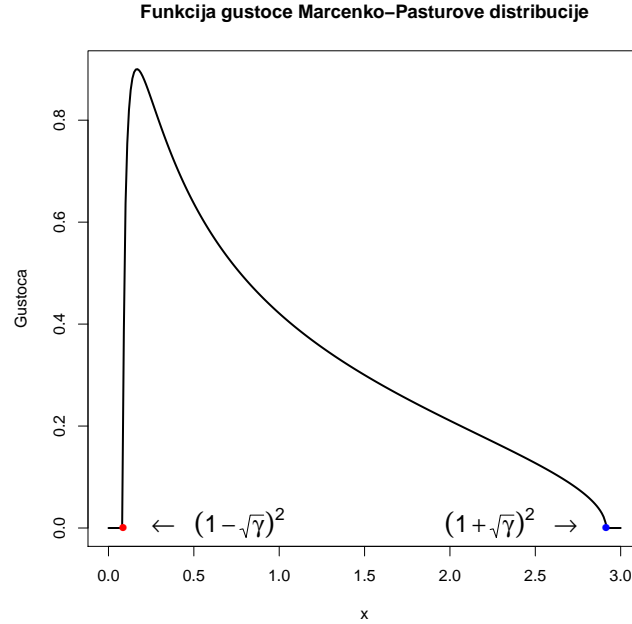
Graničnu spektralnu distribuciju normalizirane uzoračke kovarijacijske matrice otkrili su Marčenko i Pastur.

**Teorem 1.2.2.** (*Debashis i Aue [19]*) *Pretpostavimo da  $\mathbf{X}$  ima nezavisne i jednako distribuirane komponente s konačnim očekivanjem i varijancom. Ako je zadovoljen uvjet (1.2), tada s vjerojatnošću 1 niz  $(\mathbf{F}_{n^{-1}\mathbf{X}\mathbf{X}^t})$  konvergira prema neslučajnoj, tzv. Marčenko-Pasturovoj distribuciji  $F_\gamma$ . Ako je  $\gamma \in (0, 1]$ , tada  $F_\gamma$  ima gustoću*

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma} \sqrt{(b-x)(x-a)}, & \text{ako } a \leq x \leq b, \\ 0, & \text{inače,} \end{cases} \quad (1.3)$$

gdje  $a = (1 - \sqrt{\gamma})^2$  i  $b = (1 + \sqrt{\gamma})^2$ . Ako je  $\gamma > 1$ , Marčenko-Pasturova distribucija je kombinacija centrirane mase u 0 i funkcije gustoće  $f_{1/\gamma}$  s težinama  $1 - 1/\gamma$  i  $1/\gamma$ , respektivno.

Marčenko-Pasturov zakon razdiobe opisuje globalno ponašanje svojstvenih vrijednosti od  $\mathbf{X}\mathbf{X}^t$ , te je teoremom 1.2.2 kvantificirano kako su raspršene oko svoje sredine 1. Važno je uočiti da nosač razdiobe raste kako raste  $\gamma$  od 0 do  $\infty$ . Na slici 1.1 prikazana je funkcija gustoće Marčenko-Pasturove distribucije za  $\gamma = \frac{1}{2}$ . Također, naznačeni su  $a = (1 - \sqrt{\gamma})^2$  i  $b = (1 + \sqrt{\gamma})^2$  za  $\gamma = \frac{1}{2}$ .



Slika 1.1: Na grafu je prikazana funkcija gustoće Marčenko-Pasturove distribucije za  $\gamma = \frac{1}{2}$ .

Uočimo, ako  $p/n \rightarrow 0$ , granična spektralna distribucija postaje Diracova mjera u 1. Kada bi napravili pogodnu transformaciju uzoračke kovarijacijske matrice, mogli bi jasno okarakterizirati zakon distribucije polukruga koji je definiran u (1.19) kao nedegeneriranu spektralnu distribuciju u ovom slučaju.

Konačnost varijance i zahtjev da  $p$  i  $n$  rastu u beskonačnost istom brzinom su ključne pretpostavke u teoremu 1.2.2. No, teoremom 2.8. u [2] pokazano je da zaključci teorema vrijede i uz blaže pretpostavke, tj. u slučaju ako su komponente nezavisne, imaju zajedničko očekivanje te ako zadovoljavaju uvjet Lindbergovog tipa koji glasi

$$\lim_{n \rightarrow \infty} \frac{1}{\delta^2 n p} \sum_{i,t} \mathbb{E}[X_{it}^2 \mathbf{1}_{\{|X_{it}| > \delta \sqrt{n}\}}] = 0, \quad \text{za svaki } \delta > 0.$$

Teorem 1.2.2 govori o konvergenciji niza  $(F_{n^{-1} \mathbf{X} \mathbf{X}^t})$  u slučaju kada su komponente nezavisne i jednako distribuirane te konačne varijance. Razmotrimo sada slučaj kada i dalje imamo nezavisne i jednako distribuirane komponente ali ne i konačnost varijance, za to su nam potrebne sljedeće definicije.

**Definicija 1.2.3.** Funkcija  $L : (0, \infty) \rightarrow (0, \infty)$  je sporo varirajuća ako za nju vrijedi

$$\lim_{x \rightarrow \infty} \frac{L(cx)}{L(x)} = 1, \quad \text{za sve } c > 0. \quad (1.4)$$

**Definicija 1.2.4.** Kažemo da su slučajna varijabla i pripadna distribucija regularno varirajući sa indeksom  $\alpha > 0$  ako

$$\mathbb{P}(X > x) \sim p_+ \frac{L(x)}{x^\alpha} \quad i \quad \mathbb{P}(X < -x) \sim p_- \frac{L(x)}{x^\alpha}, \quad x \rightarrow \infty, \quad (1.5)$$

gdje su  $p_\pm$  nenegativne konstante takve da  $p_+ + p_- = 1$  i  $L$  je neka sporo varirajuća funkcija.

Pretpostavimo da  $\mathbf{X}$  ima regularno varirajuće komponente s indeksom  $\alpha \in (0, 2)$ . Uz uvjet (1.2) za  $\gamma \in (0, 1]$ , Belinski je u svom radu [5] teoremom 1.10 pokazao da niz  $(F_{a_{np}^{-2}\mathbf{X}\mathbf{X}^t})$  konvergira g.s. prema neslučajnoj vjerojatnosnoj mjeri s gustoćom  $\rho_\alpha^\gamma$  koja zadovoljava

$$\rho_\alpha^\gamma(x)x^{1+\alpha/2} \rightarrow \frac{\alpha\gamma}{2(1+\gamma)}, \quad x \rightarrow \infty \quad (1.6)$$

Normalizacija  $(a_k)$  definirana je tako da

$$\mathbb{P}(|X| > a_k) \sim k^{-1}, \quad k \rightarrow \infty. \quad (1.7)$$

Korištenjem Potterovih ograda (vidi npr. na stranici 25 u [8]) može se pokazati da  $a_{np}^2/n \rightarrow \infty$ . Što se tiče eksplicitnih izraza te metoda za računanje granične spektralne distribucije u slučaju beskonačne varijance, oni nisu još dostupni u ovom trenutku.

### 1.3 Granični rezultati za ekstremne svojstvene vrijednosti

U ovom odjeljku koncentrirat ćemo se na asimptotsko ponašanje najveće i najmanje svojstvene vrijednosti  $\lambda_{(1)}$  i  $\lambda_{(p)}$ , respektivno, od  $\mathbf{X}\mathbf{X}^t$ , pod uvjetom da komponente od  $\mathbf{X}$  imaju konačan četvrti moment. Kao i prije, pretpostavljamo da su komponente matrice podataka  $\mathbf{X}$  nezavisne i jednako distribuirane s realizacijom  $X$ . Nadalje, pretpostavljamo  $p \leq n$ , u suprotnom, budući da  $\mathbf{X}\mathbf{X}^t$  ima najviše  $\min(n, p)$  svojstvenih vrijednosti različitih od nule, bi vrijedilo  $\lambda_{(p)} = 0$ .

Pod uvjetom (1.2) za  $\gamma \in (0, 1]$  iz teorema 1.2.2 slijedi

$$\liminf_{n \rightarrow \infty} \frac{\lambda_{(1)}}{n} \geq (1 + \sqrt{\gamma})^2 \quad i \quad \limsup_{n \rightarrow \infty} \frac{\lambda_{(p)}}{n} \leq (1 - \sqrt{\gamma})^2 \quad g.s. \quad (1.8)$$

Bai i Silverstein u [3] daju nužne uvjete za g.s. konvergenciju  $n^{-1}\lambda_{(1)}$  koje ćemo u nastavku koristiti. Ograničenosti najvećeg dijagonalnog elemenata s najvećom svojstvenom vrijednosti daje nam sljedeću nejednakost:

$$\frac{\lambda_{(1)}}{n} \geq \max_{i=1,2,\dots,p} \frac{1}{n} \sum_{t=1}^n X_{it}^2. \quad (1.9)$$

U slučaju kada nemamo konačan četvrti moment, tj.  $\mathbb{E}[X^4] = \infty$ , iz leme B.25 u [3] slijedi

$$\limsup_{n \rightarrow \infty} \max_{i=1,2,\dots,p} \frac{1}{n} \sum_{t=1}^n X_{it}^2 = \infty \quad g.s. \quad (1.10)$$

S druge strane, ako  $\mathbb{E}[X^4] < \infty$  i  $\mathbb{E}[X] = c \neq 0$ , tada

$$\frac{1}{\sqrt{n}} \|\mathbf{X}\|_2 \geq \frac{1}{\sqrt{n}} \|\mathbb{E}[\mathbf{X}]\|_2 - \frac{1}{\sqrt{n}} \|\mathbf{X} - \mathbb{E}[\mathbf{X}]\|_2 \geq \frac{|c|p}{\sqrt{n}} - \frac{1}{\sqrt{n}} \|\mathbf{X} - \mathbb{E}[\mathbf{X}]\|_2 \rightarrow \infty \quad g.s., \quad (1.11)$$

gdje za bilo koju matricu  $\mathbf{A}$ ,  $\|\mathbf{A}\|_2$  označava spektralnu normu matrice, tj. njezinu najveću singularnu vrijednost. Sada iz (1.10) i (1.11) vidimo da su nužni uvjeti za gotovo sigurnu konvergenciju od  $n^{-1}\lambda_{(1)}$  upravo  $\mathbb{E}[X^4] < \infty$  i  $\mathbb{E}[X] = 0$ .

U nastavku ovog pregleda, osim ako drugačije nije navedeno, pretpostavljamo  $\mathbb{E}[X^2] < \infty$  i  $\mathbb{E}[X] = 0$ .

## Uzoračke kovarijacijske matrice

Bai i drugi u svom radu [30], koji je proširenje Gemanovog rada [12], pod pretpostavkom (1.2) su pokazali da

$$\frac{\lambda_{(1)}}{n} \xrightarrow{g.s.} (1 + \sqrt{\gamma})^2, \quad n \rightarrow \infty, \quad (1.12)$$

što je i optimalan rezultat uzimajući u obzir (1.10). Uz dodatnu pretpostavku  $\gamma \in (0, 1)$ , Bai i Yin su u svom radu [4] dokazali:

$$\limsup_{n \rightarrow \infty} \|n^{-1}\mathbf{X}\mathbf{X}^t - (1 + \gamma)\mathbf{I}\|_2 \leq 2\sqrt{\gamma} \quad g.s. \quad (1.13)$$

Iz relacija (1.8) i (1.13), zbog

$$\|n^{-1}\mathbf{X}\mathbf{X}^t - (1 + \gamma)\mathbf{I}\|_2 = \max \left\{ \frac{\lambda_{(1)}}{n} - (1 + \gamma), -\frac{\lambda_{(p)}}{n} + (1 + \gamma) \right\},$$

slijedi:

$$\lim_{n \rightarrow \infty} \frac{\lambda_{(1)}}{n} = (1 + \sqrt{\gamma})^2 \quad i \quad \lim_{n \rightarrow \infty} \frac{\lambda_{(p)}}{n} = (1 - \sqrt{\gamma})^2 \quad g.s. \quad (1.14)$$

Pristup koji se bazira na (1.13) omogućuje nam da smanjimo ograničenje na najmanju svojstvenu vrijednost. No, na taj način ne možemo dobiti nikakvu informaciju o minimalnim uvjetima za egzistenciju granične vrijednosti  $\lambda_{(p)}/n$  jer metoda istovremeno tretira i  $\lambda_{(1)}$  i  $\lambda_{(p)}$ . Stoga, metoda se u najboljem slučaju može iskoristiti pod uvjetom  $\mathbb{E}[X^4] < \infty$  imajući za posljedicu gubitak uvida u minimalne uvjete za graničnu vrijednost  $\lambda_{(p)}/n$ . Na

sreću, Tikhomirov je u svom članku [26] dokazao da, pod uvjetom  $\mathbb{E}[X^2] = 1$  i uz to da viši momenti ne moraju postojati, g.s. granična vrijednost od  $\lambda_{(p)}/n$  je dana s (1.14).

Uz pogodne pretpostavke na momente,  $\lambda_{(1)}$  i  $\lambda_{(p)}$  imaju *Tracy-Widom* fluktuacije oko svojih g.s. graničnih vrijednosti. Za primjer možemo uzeti Johnstoneov rad [16], u kojem je pod pretpostavkom da su komponente nezavisne i jednako distribuirane standardne normalne korištenjem pogodne varijante *Centralnog graničnog teorema* pokazao:

$$n^{2/3} \frac{(\sqrt{\gamma})^{1/3}}{(1 + \sqrt{\gamma})^{4/3}} \left( \frac{\lambda_{(1)}}{n} - \left( 1 + \sqrt{\frac{p}{n}} \right)^2 \right) \xrightarrow{d} \xi,$$

gdje granična slučajna varijabla ima *Tracy-Widom distribuciju* reda 1. Funkcija distribucije  $G_1$  te granične slučajne varijable dana sljedećom formulom

$$G_1(s) = \exp \left\{ -\frac{1}{2} \int_s^\infty [q(x) + (x-s)q^2(x)] dx \right\},$$

gdje je  $q(x)$  jedinstveno rješenje takozvane deferencijalne jednadžbe Pauliné II

$$q''(x) = xq(x) + 2q^3(x),$$

gdje  $q(x) \sim Ai(x)$  kako  $x \rightarrow \infty$ , a  $Ai(\cdot)$  je tzv. Airyjeva jezgra (vidi [27] za detalje).

## Uzoračke korelacijske matrice

O uređenim svojstvenim vrijednostima

$$\mu_{(1)} \geq \mu_{(2)} \geq \cdots \geq \mu_{(p)}$$

uzoračke korelacijske matrice  $\mathbf{R} = \mathbf{Y}\mathbf{Y}^t$  čije su komponente

$$R_{ij} = \sum_{t=1}^n \frac{X_{it}X_{jt}}{\sqrt{D_i}\sqrt{D_j}} = \sum_{t=1}^n Y_{it}Y_{jt}, \quad i, j = 1, 2, \dots, p, \quad (1.15)$$

puno manje je poznato nego o svojstvenim vrijednostima uzoračke kovarijacijske matrice  $\mathbf{X}\mathbf{X}^t$ .

Navodimo teorem bez dokaza koji ćemo u nastavku koristiti, a dokaz se može pronaći u [7].

**Teorem 1.3.1.** (*Weylova nejednakost*) Za hermitske matrice  $A$ ,  $B$  i svojstvene vrijednosti numerirane u padajuće nizove vrijedi

$$\lambda_j(A + B) \leq \lambda_i(A) + \lambda_{j-i+1}(B), \quad \text{za } i \leq j,$$

$$\lambda_j(A + B) \geq \lambda_i(A) + \lambda_{j-i+p}(B), \quad \text{za } i \geq j.$$

**Napomena 1.3.2.** U nastavku ćemo koristiti posljedicu Weylove nejednakosti koju ćemo također zvati Weylova nejednakost.

**Korolar 1.3.3.** Za hermitske matrice  $A$  i  $B$  vrijedi

$$\max_{i=1,2,\dots,p} |\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_2. \quad (1.16)$$

Često korištena notacija u nastavku će biti  $\mathbf{Y} = (Y_{it}) = (X_{it} / \sqrt{D_i})$  i

$$D_i = D_i^{(n)} = \sum_{t=1}^n X_{it}^2, \quad i = 1, 2, \dots, p; n \geq 1. \quad (1.17)$$

Ako uvedemo oznaku  $\mathbf{F} = \text{diag}(1/\sqrt{D_1}, \dots, 1/\sqrt{D_p})$ , uzoračku korelacijsku matricu možemo zapisati kao  $\mathbf{R} = \mathbf{F}^{1/2} \mathbf{X} \mathbf{X}^t \mathbf{F}^{1/2}$ , čije su svojstvene vrijednosti iste onima matrice  $\mathbf{X} \mathbf{X}^t \mathbf{F}$ . Weylova nejednakost (1.16) tada povlači

$$\begin{aligned} \max_{i=1,2,\dots,p} |\mu_{(i)} - n^{-1} \lambda_{(i)}| &\leq \|\mathbf{X} \mathbf{X}^t \mathbf{F} - n^{-1} \mathbf{X} \mathbf{X}^t\|_2 \\ &\leq n^{-1} \|\mathbf{X} \mathbf{X}^t\|_2 \|n \mathbf{F} - \mathbf{I}\|_2 \\ &= n^{-1} \lambda_{(1)} \max_{i=1,2,\dots,p} \left| \frac{n}{D_i} - 1 \right|, \end{aligned} \quad (1.18)$$

što konvergira prema 0 g.s., ako je zadovoljen uvjet  $\mathbb{E}[X^4] < \infty$ . Ekvivalenciju

$$\mathbb{E}[X^4] < \infty \iff \max_{i=1,2,\dots,p} \left| \frac{n}{D_i} - 1 \right| \xrightarrow{g.s.} 0$$

je pokazao Bai u svom radu [4].

Vođeni opisanim pristupom, uz uvjet (1.2) za  $\gamma \in (0, 1]$ , Jiang u svom radu [15], te Xiao i Zhou u [29] pokazali su sljedeći rezultat:

$$\mu_{(1)} \rightarrow (1 + \sqrt{\gamma})^2 \quad i \quad \mu_{(p)} \rightarrow (1 - \sqrt{\gamma})^2 \quad g.s.$$

## 1.4 Svojstveni vektori

Svojstveni vektori slučajnih matrica velikih dimenzija imaju važnu ulogu u modernoj statistici, fizici i drugim područjima u kojima se javlja potreba za obradom velike količine podataka. Mnoga svojstva i informacije o matricama, kao i grafovima, sadržana su u svojstvenim vektorima, pa se u mnogim popularnim algoritmima kao što su spektralno klasteriranje, metoda glavnih komponenti, te mnogim algoritmima koji se koriste u obradi signala koristi upravo dekompozicija matrice na svojstvene vrijednosti i pripadne svojstvene vektore.

Ako su komponente matrice podataka  $\mathbf{X}$  nezavisno jednako distribuirane i standardne normalne, tada je  $\mathbf{XX}^t$  *Wishartova matrica*, o čijim se svojstvenim vektorima zna mnogo. S obzirom da je  $\mathbf{XX}^t$  invarijantna na ortogonalne transformacije, ako se iskoristi pravilna ortogonalizacija, matrica svojstvenih vektora ima *Haarovu* distribuciju, tj. njezina distribucija je uniformna u prostoru ortogonalnih  $p \times p$  matrica. Ovaj rezultat proširen je na širu klasu matrica  $\mathbf{X}$  direktnom usporedbom s Wishartovom matricom koristeći takozvane univerzalne rezultate koji zahtijevaju sličnost u nekom smislu sa standardnom normalnom distribucijom. Poznato je također da matrica svojstvenih vektora ima asimptotski Haarovu distribuciju kako  $p/n \rightarrow \gamma \in (0, \infty)$ , ako se prva četiri momenta nezavisno i jednako distribuiranih komponenti podudaraju s prva četiri momenta standardne normalne distribucije.

Većina dosadašnjih istraživanja o svojstvenim vektorima slučajnih matrica provedena je na Wignerovoj matrici  $\mathbf{W} = (W_{ij})$ , koja je simerična, realna,  $n \times n$  matrica čije su komponente slučajne varijable  $W_{ij}$ ,  $1 \leq i \leq j \leq n$ . Slučajne varijable, komponente Wignerove matrice su nezavisne i jednako distribuirane te nultog očekivanja i jedinične varijance. Uočimo, za  $p = n$  i istu prigodnu distribuciju komponenti uzoračka kovarijacijska matrica i kvadrat Wignerove matrice  $\mathbf{W}^2 = \mathbf{WW}^t$  imaju slična spektralna svojstva. Detaljnijom analizom može se vidjeti da imaju istu strukturu do na simetričnost matrice  $\mathbf{W}$ .

Bai je u svom radu [31] pokazao da je granična spektralna distribucija od  $(\mathbf{W}/\sqrt{n})$  razdioba polukruga  $G$ . Gustoća razdiobe polukruga definira se s

$$g(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{\{|x| \leq 2\}}. \quad (1.19)$$

Distribucija polukruga i Marčenko-Pasturova distribucija  $F_\gamma$ , za  $\gamma = 1$ , su povezane na sljedeći način: ako  $Y \sim G$  tada  $Y^2 \sim F_1$ .

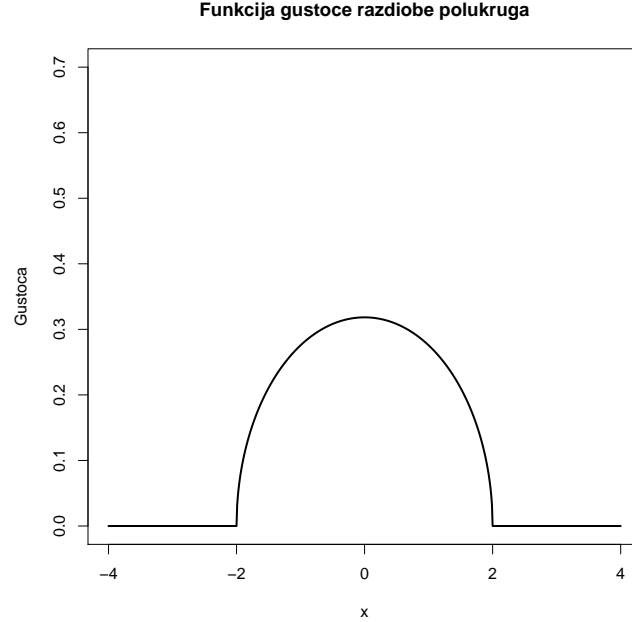
Na slici 1.2 prikazana je funkcija gustoće razdiobe polukruga. Iz slike je očita povezanost s nazivom distribucije.

U svrhu pregleda nekih svojstava svojstvenih vektora koncentrirat ćemo se na svojstvene vektore matrice  $\mathbf{W}$  budući da  $\mathbf{W}^2$  i  $\mathbf{W}$  imaju iste svojstvene vektore. Promatramo svojstvene vektore  $\mathbf{v}_i$ ,  $i = 1, \dots, n$ , gdje će nam  $\mathbf{v}_i$  biti oznaka za svojstveni vektor koji pripada  $i$ -toj po redu najvećoj svojstvenoj vrijednosti, dodatno pretpostavljamo da su vektori jedinični, tj.  $\|\mathbf{v}_i\|_{\ell_2} = 1$  u  $\ell_2$  normi i da je prvi element vektora  $\mathbf{v}_i$  koji je različit od nule pozitivan.

U literaturi se često pretpostavlja da komponente imaju *subeksponecijalnu* distribuciju. U [17] Mikosch koristi sljedeću definiciju. Za pozitivnu slučajnu varijablu  $X$  s neograničenim nosačem i njezinu distribuciju kažemo da su subeksponecijalni ako za niz  $(X_i)$  nezavisnih i jednako distribuiranih slučajnih varijabli s istom distribucijom kao i  $X$  vrijedi:

$$\text{Za sve } n \geq 2 : \mathbb{P}(S_n > x) = \mathbb{P}(M_{n>x})(1 + o(1)), \quad \text{kad } x \rightarrow \infty,$$

gdje je  $S_n = X_1 + \dots + X_n$  i  $M_n = \max(X_1, \dots, X_n)$ .



Slika 1.2: Na grafu je prikazana funkcija gustoće za distribuciju polukruga.

U [18] O'Rourke koristi sljedeću definiciju. Za slučajnu varijablu  $W$  kažemo da ima subekspencijalnu distribuciju s eksponentom  $\alpha > 0$  ako postoji konstanta  $\beta > 0$  takva da

$$\mathbb{P}(|W| > x) \leq \beta \exp(-x^\alpha / \beta), \quad x > 0.$$

Pretpostavku da Wignerova matrica ima subekspencijalne komponente s parametrom  $\alpha$  koristio je i O'Rourke u [18], te su ovo rezultati nekih od teorema iz njegovog rada. Postoji konstanta  $C_\alpha > 0$  takva da je vjerojatnost da je spektar matrice  $\mathbf{W}$  jednostavan, tj. da su sve nultočke karakterističnog polinoma jednostruke i da su sve koordinate svakog vektora  $\mathbf{v}_i$  nenegativne je barem  $1 - C_\alpha n^{-\alpha}$ . Po notaciji i konvenciji koju koristimo, svojstveni vektori su jedinstveni s velikom vjerojatnošću, tj. preciznije, ako je  $W_{11}$  standardna normalna, tada je  $\mathbf{v}_i$  uniformno distribuiran na

$$S_+^{n-1} := \{\mathbf{x} = (x_1, \dots, x_n) : \|\mathbf{x}\|_{\ell_2} = 1 \text{ i } x_1 > 0\}.$$

Svojstveni vektor čije komponente su distribucije lakog repa (npr. subekspencijalne) ponaša se kao slučajni vektor uniformno distribuiran na  $S_+^{n-1}$ .

Ako je  $\mathbf{v} = (v_1, v_2, \dots, v_n)^t$  slučajni vektor uniformno distribuiran na  $S_+^{n-1}$ , vjerojatnosna ograničenja na njegove koordinate je pokazao također O'Rourke u gornjem radu i ona su:

$$v_{\max} := \max_{i=1,2,\dots,n} |v_i| \leq C \sqrt{\frac{\log n}{n}} \quad \text{ i } \quad v_{\min} := \min_{i=1,2,\dots,n} |v_i| \geq \frac{c}{n^{3/2}},$$



s vjerojatnošću  $1 - o(1)$  za svaki  $C > 1$  i  $c \in [0, 1)$ .

Važno je napomenuti da u slučaju ako  $\mathbf{W}$  ima teške repove ponašanje svojstvenih vektora je potpuno drugačije, te su svojstveni vektori matrica  $\mathbf{W}$  i  $\mathbf{X}\mathbf{X}^t$  jako različiti.

## 1.5 Teški repovi

Distribucije kod kojih određeni moment nisu konačni katkad zovemo distribucijama teških repova. Kao što smo vidjeli do sada, za konvergenciju po distribuciji empirijske spektralne distribucije i najveće svojstvene vrijednosti uzoračke kovarijacijske matrice prema Marčenko-Pasturovoj i Tracy-Widomovoj distribuciji nužni su bili konačni prvi i drugi moment komponenata matrice podataka.

U (1.10) i (1.11) vidjeli smo kako se ponašanje najveće svojstvene vrijednosti  $\lambda_{(1)}$  drastično promijeni ovisno o slučaju  $\mathbb{E}[X^4] < \infty$  ili  $\mathbb{E}[X^4] = \infty$ .

Za daljnju analizu graničnih rezultata u klasičnoj teoriji o procesima parcijalnih suma i maksimuma potrebno nam je više nego beskonačan četvrti moment. Stoga, pretpostavljamo uvjet regularne varijacije (1.5) na rep matrice  $\mathbf{X}$ , za neki  $\alpha \in (0, 4)$ , nenegativne konstante  $p_{\pm}$  takve da je  $p_{+} + p_{-} = 1$  i sporo varijirajuću funkciju  $L$  koja zadovoljava (1.4). Također, koristit ćemo se pojmovima  $X$  regularno varirajuća slučajna varijabla,  $\mathbf{X}$  regularno varirajuća slučajna matrica i sličnima. Nadalje, koristimo i normalizaciju svojstvenih vrijednosti  $\lambda_{(i)}$  s  $(a_{np}^2)$  koja je definirana u (1.7).

Svojstvene vrijednosti matrice  $\mathbf{X}\mathbf{X}^t$  teških repova je prve počeo proučavati Soshnikov [23, 22]. On je pod (1.2) i (1.5) za  $\alpha \in (0, 2)$  pokazao da

$$\frac{\lambda_{(1)}}{a_{np}^2} \xrightarrow{d} \zeta, \quad n \rightarrow \infty, \quad (1.20)$$

gdje  $\zeta$  prati *Fréchetovu distribuciju* s parametrom  $\alpha/2$ :

$$\Phi_{\alpha/2}(x) = e^{-x^{-\alpha/2}}, \quad x > 0. \quad (1.21)$$

Kasnije je pokazano (1.20) i za  $\alpha \in [2, 4)$ , uz dodatnu pretpostavku na centriranost komponenti matrice podataka. Kasnije su Soshnikov i ostali dokazali i konvergenciju točkovnih procesa normaliziranih svojstvenih vrijednosti, od kuda je lako slijedila zajednička granična distribucija  $k$  najvećih svojstvenih vrijednosti. Davis i suradnici su u [10, 11] proširili ove rezultate dopuštajući generalnije stope rasta  $p$  nego što je određeno uvjetom (1.2) i dopuštajući linearnu zavisnost između redaka i stupaca matrice  $\mathbf{X}$ .

## 1.6 Konvergencija po distribuciji slučajnih mjera

U ovom odjeljku uvodimo koncept konvergencije po distribuciji točkovnih procesa i slučajnih mjera koji će biti korišteni na više mjesta u nastavku.

Neka je  $\mathbb{S}$  potpun (tj. takav da svaki Cauchyev niz konvergira u njemu) i separabilan (postoji prebrojiv gust podskup) metrički prostor s metrikom  $d$  i neka je  $\mathcal{S}$  Borelova  $\sigma$ -algebra podskupova od  $\mathbb{S}$  generirana otvorenim skupovima. Pretpostavimo da je  $(\Omega, \mathcal{A}, \mathbb{P})$  vjerojatnosni prostor. Slučajni element  $X$  iz  $\mathcal{S}$  je izmjerivo preslikavanje iz prostora  $(\Omega, \mathcal{A})$  u  $(\mathbb{S}, \mathcal{S})$ .

Ako se radi o slučajnoj varijabli tada se događaj  $\omega \in \Omega$  preslikava u broj iz skupa  $\mathbb{R}$ . Općenito, ako se radi o slučajnom elementu, tada se događaj  $\omega \in \Omega$  preslikava u element metričkog prostora  $\mathbb{S}$ . Prostori i pripadni slučajni elementi koji se najčešće pojavljuju u literaturi dani su u sljedećoj tablici:

Metrički prostor $\mathbb{S}$	Slučajni element $X$
$\mathbb{R}$	Slučajna varijabla
$\mathbb{R}^d$	Slučajni vektor
$\mathbb{R}^\infty$	Slučajni niz
$C[0, \infty)$ prostor neprekidnih realnih funkcija s vrijednostima u $[0, \infty)$	Slučajni proces s neprekidnim putevima
$M_p(\mathbb{S})$ , prostor točkovnih mjera na prostoru $\mathbb{S}$	Stohastički točkovni proces na $\mathbb{S}$
$M_+(\mathbb{S})$ , prostor Radonovih mjera na lijepom prostoru $\mathbb{S}$	Radonova mjera na $\mathbb{S}$

Za dani niz slučajnih elemenata  $\{X_n, n \geq 0\}$  iz  $\mathbb{S}$  postoji odgovarajući niz distribucija u  $\mathcal{S}$

$$P_n = \mathbb{P} \circ X_n^{-1} = \mathbb{P}(X_n \in \cdot), \quad n \geq 0.$$

Vjerojatnosnu mjeru  $P_n$  zovemo distribucijom od  $X_n$ . Kažemo da  $X_n$  konvergira po distribuciji prema  $X_0$  (pišemo  $X_n \xrightarrow{d} X_0$ ) ako za svaku  $f \in C(\mathbb{S})$ , klase ograničenih, neprekidnih i realnih funkcija vrijedi

$$\mathbb{E}f(X_n) = \int_{\mathbb{S}} f(x) P_n(dx) \rightarrow \mathbb{E}f(X_0) = \int_{\mathbb{S}} f(x) P_0(dx)$$

Ovakva definicija konvergencije, u terminima test funkcija  $f \in C(\mathbb{S})$  je vrlo fleksibilna pa stoga i vrlo primjenjiva te dobro definirana za bilo koji metrički prostor  $\mathbb{S}$ .

## Radonova mjera i točkovne mjere

Pretpostavimo da je  $\mathbb{S}$  zgodan prostor. *Zgodan* prostor znači da bi  $\mathbb{S}$  trebao biti lokalno kompaktan topološki prostor s prebrojivom bazom,  $\mathbb{S}$  možemo zamišljati kao konačnodimenzionalni Euklidski prostor ili  $\mathbb{R}^d$ . Pretpostavimo da uz  $\mathbb{S}$  dolazi i pripadna  $\sigma$ -algebra  $\mathcal{S}$  koja može biti generirana otvorenim skupovima ili pravokutnicima iz  $\mathbb{S}$ .

Postoji više načina za modeliranje slučajne distribucije točaka iz  $\mathbb{S}$ . U nastavku je prikazan način koji je preferiran u literaturi (Resnick [21, 20]).

Mjera  $\mu : \mathcal{S} \rightarrow [0, \infty]$  je preslikavanje koje pridružuje pozitivne vrijednosti skupovima iz  $\mathcal{S}$  takvo da

1.  $\mu(\emptyset) = 0$  i  $\mu(A) \geq 0$  za svaki  $A \in \mathcal{S}$
2. ako su  $\{A_n, n \geq 1\}$  međusobno disjunktni skupovi iz  $\mathcal{S}$ , tada vrijedi  $\sigma$ -aditivnost

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Za mjeru  $\mu$  kažemo da je Radonova ako  $\mu(K) < \infty$  za svaki kompaktan podskup od  $\mathbb{S}$ . Definiramo

$$M_+(\mathbb{S}) = \{\mu : \mu \text{ je nenegativna i } \mu \text{ je Radonova}\}.$$

Prostor točkovnih mjera na prostoru  $\mathbb{S}$ , u oznaci  $M_p(\mathbb{S})$ , je podskup od  $M_+(\mathbb{S})$  takav da  $\mu(A) \in \mathbb{Z}_+$ , za  $A \subseteq \mathbb{S}$ . Točkovni proces je slučajni element s vrijednostima u  $M_p(\mathbb{S})$ .

Nadalje,  $M_+(\mathbb{S})$  postaje potpun i separabilan prostor ako mu pridružimo tzv. *vague* topologiju. U slučaju beskonačnih mjera iz  $M_+(\mathbb{E})$ , jer su mjere Radonove, za test funkcije koristimo funkcije koje su 0 na komplementu kompaktnog skupa. Definiramo

$$C_K^+(\mathbb{S}) = \{f : \mathbb{S} \mapsto \mathbb{R}_+ : f \text{ je neprekidna s kompaktnim nosačem}\}.$$

Konvergenciju u  $M_+(\mathbb{S})$  definiramo na sljedeći način:

Ako je  $\mu_n \in M_+(\mathbb{S})$  za  $n \geq 0$ , tada  $\mu_n$  konvergira (*vague* konvergencija) prema  $\mu_0$ , u oznaci  $\mu_n \xrightarrow{v} \mu$ , ako za sve  $f \in C_K^+(\mathbb{S})$  vrijedi

$$\mu_n(f) := \int_{\mathbb{S}} f(x) \mu_n(dx) \rightarrow \mu_0(f) := \int_{\mathbb{S}} f(x) \mu_0(dx), \quad \text{kad } n \rightarrow \infty.$$

Ispostavljena *vague* konvergencija definira na  $\mathbb{S}$  i metrizabilnu topologiju, pa možemo definirati konvergenciju po distribuciji slučajnih elemenata u  $M_+(\mathbb{S})$ .

Rezultati vezani uz konvergenciju po distribuciji točkovnih procesa i slučajnih mjera mogu se pronaći u Resnickovim knjigama [21] i [20]. Također, vrlo intuitivan opis ove konvergencije uz mnoge primjere i primjene može se pronaći u Mikoschevoj knjizi [17].

## Poglavlje 2

# Svojstvene vrijednosti kovarijacijske matrice

U ovom poglavlju bavit ćemo se asimptotskim svojstvima najveće *svojstvene vrijednosti* i odgovarajućeg *svojstvenog vektora* uzoračkih kovarijacijskih matrica velikih dimenzija za vremenske nizove teških repova s nezavisno jednako distribuiranim komponentama. Posebnu pažnju posvećujemo slučaju kada dimenzije  $p$  (broj redaka) i  $n$  (broj stupaca) matrice podataka rastu u beskonačnost, no ne nužno istom brzinom.

### Stope rasta od $p$

U mnogim primjenama slučajnih matrica velikih dimenzija nije realna pretpostavka da dimenzija podataka  $p$  i količina podataka  $n$  rastu istom brzinom, tj. u praksi je malo vjerojatno da će uvjet (1.2) biti zadovoljen. Stoga Heiny i Mikosch u [13] uvode stope rasta od  $(p_n)$  koje se nisu pojavljivale u literaturi:

$$p = p_n = n^\beta \ell(n), \quad n \geq 1, \quad (2.1)$$

gdje je  $\ell$  sporo varirajuća funkcija koja zadovoljava (1.4) i  $\beta \geq 0$ . Ako je  $\beta = 0$ , pretpostavljamo da  $\ell(n) \rightarrow \infty$ .

### Notacija

Neka je

$$\mathbf{Z} = \mathbf{Z}_n = (Z_{it})_{i=1,2,\dots,p; t=1,2,\dots,n}$$

matrica podataka dimenzije  $p \times n$  čiji su stupci obzervacije  $p$ -dimenzionalnih vremenskih nizova. Nadalje, pretpostavljamo da su komponente matrice  $\mathbf{Z}$  nezavisne i jednako distribuirane te da zadovoljavaju uvjet regularne varijacije (1.5) za neki  $\alpha \in (0, 4)$ . Svojstvene

vrijednosti uređene u padajući poredak matrice  $\mathbf{Z}\mathbf{Z}^t$  označimo s

$$\lambda_{(1)}, \lambda_{(2)}, \dots, \lambda_{(p)}. \quad (2.2)$$

Od velike važnosti bit će nam veličine  $(Z_{it}^2)_{i=1,2,\dots,p;t=1,2,\dots,n}$ , kao i pridružene uređajne statistike

$$Z_{(1),np}^2 \geq Z_{(2),np}^2 \geq \dots Z_{(np),np}^2 \quad n, p \geq 1. \quad (2.3)$$

Također, jako važne će nam biti i sume redaka

$$D_i^{\rightarrow} = D_i^{(n),\rightarrow} = \sum_{t=1}^n Z_{it}^2, \quad i = 1, 2, \dots, p; \quad n = 1, 2, \dots, \quad (2.4)$$

koje ćemo kraće označavati s  $D^{\rightarrow}$  i njihove uređene vrijednosti

$$D_{(1)}^{\rightarrow} = D_{L_1}^{\rightarrow} \geq \dots \geq D_{(p)}^{\rightarrow} = D_{L_p}^{\rightarrow}, \quad (2.5)$$

gdje b.s.o.m.p. da je  $(L_1, L_2, \dots, L_p)$  odgovarajuća permutacija od  $(1, 2, \dots, p)$  za fiksni  $n$ .

Jednako važne kao sume redaka će nam biti i sume stupaca

$$D_t^{\downarrow} = D_t^{(n),\downarrow} = \sum_{i=1}^p Z_{it}^2, \quad t = 1, 2, \dots, n; \quad p = 1, 2, \dots, \quad (2.6)$$

koje ćemo kreće označavati s  $D^{\downarrow}$ . Za njihove uređene vrijednosti koristit ćemo notaciju analognu onoj za sume redaka iz (2.5).

Sada, uvodimo još notaciju vezanu uz norme vektora i matrica. Euklidsku normu vektora  $\mathbf{v}$  dimenzije  $p$  označavat ćemo s  $\|\mathbf{v}\|_{\ell_2}$ . Za kvadratnu matricu  $\mathbf{C}$  dimenzije  $p \times p$  s  $\lambda_i(\mathbf{C})$  označimo njene singularne vrijednosti, a pridružene uređajne statistike s

$$\lambda_{(1)}(\mathbf{C}) \geq \lambda_{(2)}(\mathbf{C}) \geq \dots \lambda_{(p)}(\mathbf{C}).$$

Nadalje, za proizvoljnu  $p \times n$  matricu  $\mathbf{A}$  koristit ćemo još i sljedeće norme: *spektralnu normu*  $\|\mathbf{A}\|_2 = \sqrt{\lambda_{(1)}(\mathbf{A}\mathbf{A}^t)}$ , *Frobeniusovu normu*  $\|\mathbf{A}\|_F = (\sum_{i=1}^p \sum_{j=1}^n |a_{ij}|^2)^{1/2}$  te  *$\infty$ -normu*  $\|\mathbf{A}\|_{\infty} = \max_{i=1,2,\dots,p} \sum_{j=1}^n |a_{ij}|$ .

## 2.1 Temeljne aproksimacije

Ovo poglavlje započinjemo teoremom kojim Heiny i Mikosch u [13] daju temeljne aproksimacije za svojstvene vrijednosti matrice  $\mathbf{Z}\mathbf{Z}^t$ . Veličine koje aproksimiraju tražene vrijednosti su jednostavne strukture te imaju vrlo korisna svojstva. Jedno od njih je da je njihovo asimptotsko ponašanje naslijeđeno od svojstvenih vrijednosti te ima utjecaja i na svojstvene vektore.

**Teorem 2.1.1.** *Neka je  $Z$   $p \times n$  matrica s nezavisno jednako distribuiranim komponentama. Pretpostavljamo da su zadovoljeni sljedeći uvjeti:*

- *uvjet regularne varijacije (1.5) za neki  $\alpha \in (0, 4)$ ,*
- *$\mathbb{E}[Z] = 0$  za  $\alpha \geq 2$ ,*
- *niz prirodnih brojeva  $(p_n)$  raste kao u (2.1) za neki  $\beta \geq 0$ .*

*Tada vrijede sljedeći rezultati:*

1. *Ako je  $\beta \in [0, 1]$ , tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)} - D_{(i)}^{\rightarrow}| \xrightarrow{\mathbb{P}} 0. \quad (2.7)$$

2. *Ako je  $\beta > 1$ , tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,n} |\lambda_{(i)} - D_{(i)}^{\downarrow}| \xrightarrow{\mathbb{P}} 0. \quad (2.8)$$

3. *Ako je  $\min(\beta, \beta^{-1}) \in ((\alpha/2 - 1)_+, 1]$ , tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)} - Z_{(i),np}^2| \xrightarrow{\mathbb{P}} 0. \quad (2.9)$$

**Napomena 2.1.2.** *Uočimo da se za razliku od slučajeva (1) i (3), u slučaju (2) traži maksimalna vrijednosti po skupu indeksa  $\{1, 2, \dots, n\}$ . Budući da  $p \times p$  matrica  $ZZ^t$  i  $n \times n$  matrica  $Z^tZ$  imaju iste pozitivne svojstvene vrijednosti, vrijedi  $\lambda_{(i)} = 0$  za  $i = \min(p, n) + 1, \dots, \max(p, n)$ . Nadalje, u slučaju kada je  $\beta \in (0, 1)$  za dovoljno veliki  $n$  je  $\min(p, n) = p$ , dok je za  $\beta > 1$   $\max(p, n) = n$  za dovoljno veliki  $n$ . Dakle, jedino za  $\beta = 1$  su moguća oba slučaja  $n \leq p$  i  $p \leq n$ .*

**Napomena 2.1.3.** *Važno je uočiti da uvjet  $\min(\beta, \beta^{-1}) \in ((\alpha/2 - 1)_+, 1]$ , u slučaju (3), predstavlja restrikciju za  $\alpha \in (2, 4)$ . Također, iz ovog uvjeta slijedi  $(\min(p, n))a_{np}^{-2} \rightarrow 0$ . Gornja implikacija zapravo govori da se centriranje  $a_{np}^{-2}D_i^{\rightarrow}$  i  $a_{np}^{-2}D_i^{\downarrow}$  u graničnim teoremima može izbjeći.*

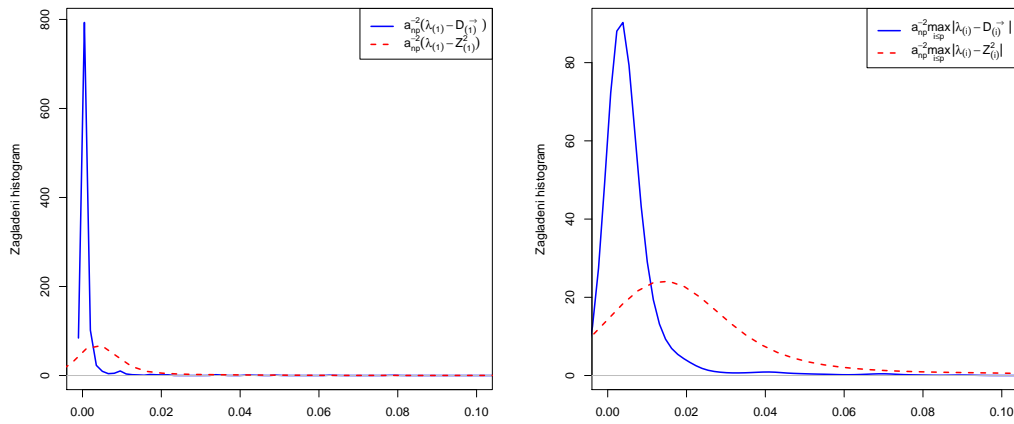
**Napomena 2.1.4.** *Na slici 2.1 ilustriramo razlike u aproksimacijama, tj. uspoređujemo distribucije pogrešaka kada svojstvene vrijednosti  $(\lambda_{(i)})$  aproksimiramo s  $(D_{(i)}^{\rightarrow})$  kao što je sugerirano u prethodnom teoremu s (2.7), te sa  $(Z_{(i),np}^2)$  kao što je sugerirano u (2.9). U primjeru, gustoća  $Z$  za svaku komponentu matrice je dana s*

$$f_Z(x) = \begin{cases} \frac{\alpha}{(4|x|)^{\alpha+1}}, & \text{ako } |x| > 1/4 \\ 1, & \text{inače.} \end{cases} \quad (2.10)$$

Na lijevoj slici pažnju obraćamo na najveću svojstvenu vrijednost  $\lambda_{(1)}$ . Prikazani su zaglađeni histogrami pogrešaka aproksimacije  $a_{np}^{-2}(\lambda_{(1)} - D_{(1)}^{\rightarrow})$  i  $a_{np}^{-2}(\lambda_{(1)} - Z_{(1),np}^2)$  za koje se (vidi lemu 22 u [25]) pretpostavlja da su nenegativne.

Na desnoj slici pak gledamo zaglađene histograme za maksimalne vrijednosti pogrešaka  $a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)} - D_{(i)}^{\rightarrow}|$  i  $a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)} - Z_{(i),np}^2|$  kao što je sugerirano u (2.7) i (2.9).

Iz grafova na slikama možemo zaključiti da su aproksimacije s  $(D^{\rightarrow(i)})$  mnogo bolje od onih sa  $(Z_{(1),np}^2)$ .



Slika 2.1: Zaglađeni histogrami pogrešaka aproksimacije za normalizirane svojstvene vrijednosti  $(a_{np}^{-2}\lambda_{(i)})$  za komponente  $Z_{it}$  s gustoćom (2.10), uz  $\alpha = 1.6$ ,  $\beta = 1$ ,  $n = 1000$  i  $p = 200$ .

Dokaz teorema je vrlo zahtjevan i dokazuje se pomoću niza lema koje se mogu pronaći, kao i sam dokaz teorema u [13]. Veliki korak u dokazu čini i sljedeći teorem, čiji se dokaz može također pronaći u radu [13] Heinyja i Mikoscha.

**Teorem 2.1.5.** *Pretpostavimo da vrijede uvjeti teorema 2.1.1 za matricu podataka  $\mathbf{Z}$  i stope rasta  $(p_n)$ .*

1. *Ako je  $\beta \in [0, 1]$ , tada*

$$a_{np}^{-2} \|\mathbf{Z}\mathbf{Z}^t - \text{diag}(\mathbf{Z}\mathbf{Z}^t)\|_2 \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

2. *Ako je  $\beta \geq 1$ , tada*

$$a_{np}^{-2} \|\mathbf{Z}^t\mathbf{Z} - \text{diag}(\mathbf{Z}^t\mathbf{Z})\|_2 \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

**Napomena 2.1.6.** Uočimo da za  $\beta \geq 1$  možemo pisati  $n = p^{1/\beta} \ell(p)$ , tj. zamjenom  $p$  i  $n$  drugi dio teorema slijedi direktno iz prvog.

Nadalje, iz teorema 2.1.5 slijedi da je najveća svojstvena vrijednost matrice  $\mathbf{Z}\mathbf{Z}^t$  određena dijagonalom matrice  $\mathbf{Z}\mathbf{Z}^t$ . Treba uzeti u obzir još da dijagonalni elementi nemaju ključnu ulogu kada se promatra Wignerova matrica teških repova.

Teorem 2.1.5 ima za posljedicu sljedeći rezultat o aproksimacijama najvećih svojstvenih vrijednosti matrica  $\mathbf{Z}\mathbf{Z}^t$  i  $\mathbf{Z}^t\mathbf{Z}$  s pripadnim dijagonalnim elementima  $\text{diag}(\mathbf{Z}\mathbf{Z}^t)$  i  $\text{diag}(\mathbf{Z}^t\mathbf{Z})$ .

**Korolar 2.1.7.** Pretpostavimo da vrijede uvjeti teorema 2.1.1 za matricu podataka  $\mathbf{Z}$  i stope rasta  $(p_n)$ .

1. Ako je  $\beta \in [0, 1]$ , tada

$$a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)} - \lambda_{(i)}(\text{diag}(\mathbf{Z}\mathbf{Z}^t))| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

2. Ako je  $\beta > 1$ , tada

$$a_{np}^{-2} \max_{i=1,2,\dots,n} |\lambda_{(i)} - \lambda_{(i)}(\text{diag}(\mathbf{Z}^t\mathbf{Z}))| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

*Dokaz.* Neka su  $\mathbf{A}$  i  $\mathbf{B}$   $p \times p$  simetrične matrice. Po Weylovoj nejednakosti (1.16) slijedi

$$\max_{i=1,2,\dots,p} |\lambda_{(i)}(\mathbf{A} + \mathbf{B}) - \lambda_{(i)}(\mathbf{A})| \leq \|\mathbf{B}\|_2.$$

Sada, ako stavimo  $(\mathbf{A} + \mathbf{B}) = \mathbf{Z}\mathbf{Z}^t$  i  $\mathbf{A} = \text{diag} \mathbf{Z}\mathbf{Z}^t$  (ili  $(\mathbf{A} + \mathbf{B}) = \mathbf{Z}^t\mathbf{Z}$  i  $\mathbf{A} = \text{diag} \mathbf{Z}^t\mathbf{Z}$ ) slijedi upravo tvrdnja korolara.  $\square$

Tvrdnje (2.7) i (2.8) su direktne posljedice prethodnog korolara ako primijetimo da vrijede sljedeće jednakosti:  $\lambda_{(i)}(\text{diag}(\mathbf{Z}\mathbf{Z}^t)) = D_{(i)}^{\rightarrow}$  i  $\lambda_{(i)}(\mathbf{Z}^t\mathbf{Z}) = D_{(i)}^{\downarrow}$ ,  $i = 1, 2, \dots, \min(p, n)$ .

## 2.2 Svojstveni vektori

Pretpostavimo da vrijede uvjeti teorema 2.1.5 i da je  $\beta \in [0, 1]$ . Kao što smo vidjeli u teoremu 2.1.5  $\text{diag}(\mathbf{Z}\mathbf{Z}^t)$  je aproksimacija za  $\mathbf{Z}\mathbf{Z}^t$  u spektralnoj normi. Jer je  $p \times p$  matrica  $\text{diag}(\mathbf{Z}\mathbf{Z}^t)$  dijagonalna, njezini jedinični svojstveni vektori su vektori kanonske baze  $\mathbf{e}_j \in \mathbb{R}^p$ ,  $j = 1, 2, \dots, p$ . Motivirani time možemo postaviti pitanje jesu li vektori  $(\mathbf{e}_j)$  dobra aproksimacija za svojstvene vektore  $(\mathbf{v}_j)$  matrice  $\mathbf{Z}\mathbf{Z}^t$ . S  $\mathbf{v}_j$  označavat ćemo svojstveni vektor  $j$ -te najveće svojstvene vrijednosti  $\lambda_{(j)}$  matrice  $\mathbf{Z}\mathbf{Z}^t$ . Pripadni jedinični svojstveni vektor  $j$ -te najveće svojstvene vrijednosti matrice  $\mathbf{Z}\mathbf{Z}^t$  je  $\mathbf{e}_{L_j}$ , gdje je  $L_j$  uređaj kao u (2.5). Slutnju da je  $\mathbf{e}_{L_j}$  dobra aproksimacija za  $\mathbf{v}_j$  dokazao je Heiny sljedećim teoremom.



**Teorem 2.2.1.** *Pretpostavimo da vrijede uvjeti teorema 2.1.1 i neka je  $\beta \in [0, 1]$ . Tada za svaki fiksni  $k \geq 1$ ,*

$$\|\mathbf{v}_k - \mathbf{e}_{L_k}\|_{\ell_2} \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

Za dokaz teorema potreban nam je sljedeći pomoćni rezultat koji navodimo bez dokaza (propozicija A.1 u [6]).

**Propozicija 2.2.2.** *Neka je  $\mathbf{H}$  Hermiteova matrica i  $\mathbf{v}$  jedinični vektor takav da za neki  $\lambda \in \mathbb{R}$  i  $\varepsilon > 0$ ,*

$$\mathbf{H}\mathbf{v} = \lambda\mathbf{v} + \varepsilon\mathbf{w},$$

*gdje je  $\mathbf{w}$  jedinični vektor takav da  $\mathbf{w} \perp \mathbf{v}$ .*

1. *Tada  $\mathbf{H}$  ima svojstvenu vrijednost  $\lambda_\varepsilon$  takvu da  $|\lambda - \lambda_\varepsilon| \leq \varepsilon$ .*
2. *Ako  $\mathbf{H}$  ima samo jednu svojstvenu vrijednost (s uračunatom višestrukosti) takvu da  $|\lambda - \lambda_\varepsilon| \leq \varepsilon$ , a sve ostale svojstvene vrijednosti su na udaljenosti  $d > \varepsilon$  od  $\lambda$ . Tada za jedinični svojstveni vektor  $\mathbf{v}_\varepsilon$  koji pripada svojstvenoj vrijednosti  $\lambda_\varepsilon$  vrijedi*

$$\|\mathbf{v}_\varepsilon - \mathbf{P}_\mathbf{v}(\mathbf{v}_\varepsilon)\|_{\ell_2} \leq \frac{2\varepsilon}{d - \varepsilon},$$

*gdje je  $\mathbf{P}_\mathbf{v}$  oznaka za ortogonalni projektor na linearnu mnogostrukost razapetu vektorom  $\mathbf{v}$ .*

Sada prelazimo na dokaz teorema 2.2.1.

*Dokaz.* Fiksirajmo  $k \geq 1$ . Ako uzmemo u obzir da  $p \rightarrow \infty$ , onda možemo pretpostaviti da je  $k \leq p$  za dovoljno velik  $n$ . Za matricu  $\mathbf{Z}\mathbf{Z}^t - \text{diag}(\mathbf{Z}\mathbf{Z}^t)$  komponente stupaca, za  $j = 1, 2, \dots, p$ , su dane sa

$$\mathbf{Z}\mathbf{Z}^t \mathbf{e}_j - D_j^\rightarrow \mathbf{e}_j = \left( \sum_{t=1}^n Z_{1t} Z_{jt}, \dots, \sum_{t=1}^n Z_{j-1,t} Z_{jt}, 0, \sum_{t=1}^n Z_{j+1,t} Z_{jt}, \dots, \sum_{t=1}^n Z_{pt} Z_{jt} \right)^t.$$

Po teoremu 2.1.5(1),

$$a_{np}^{-2} \max_{j=1,2,\dots,p} \|\mathbf{Z}\mathbf{Z}^t \mathbf{e}_j - D_j^\rightarrow \mathbf{e}_j\|_{\ell_2} \leq a_{np}^{-2} \|\mathbf{Z}\mathbf{Z}^t - \text{diag}(\mathbf{Z}\mathbf{Z}^t)\|_2 \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty. \quad (2.11)$$

Ako stavimo  $\mathbf{H}^{(n)} = a_{np}^{-2} \mathbf{Z}\mathbf{Z}^t$ ,  $\mathbf{v}^{(n)} = \mathbf{e}_{L_k} \in \mathbb{R}^p$  i  $\lambda^{(n)} = a_{np}^{-2} D_{L_k}^\rightarrow$ , tada vrijedi

$$a_{np}^{-2} \mathbf{Z}\mathbf{Z}^t \mathbf{e}_{L_k} = a_{np}^{-2} D_{L_k}^\rightarrow \mathbf{e}_{L_k} + \varepsilon^{(n)} \mathbf{w}^{(n)},$$

gdje je  $\mathbf{w}^{(n)} = \|\mathbf{Z}\mathbf{Z}^t \mathbf{e}_{L_k} - D_{L_k}^{-1}(\mathbf{Z}\mathbf{Z}^t \mathbf{e}_{L_k} - D_{L_k}^{-1} \mathbf{e}_{L_k})\|_{\ell_2}^{-1}(\mathbf{Z}\mathbf{Z}^t \mathbf{e}_{L_k} - D_{L_k}^{-1} \mathbf{e}_{L_k})$  jedinični vektor, a po (2.11)  $\varepsilon^{(n)} = a_{np}^{-2} \|\mathbf{Z}\mathbf{Z}^t \mathbf{e}_j - D_j^{-1} \mathbf{e}_j\|_{\ell_2} \xrightarrow{\mathbb{P}} 0$ .

Da bi mogli iskoristiti propoziciju 2.2.2, moramo još pokazati da s vjerojatnošću koja konvergira k 1, matrica  $\mathbf{Z}\mathbf{Z}^t$  nema drugih svojstvenih vrijednosti u dovoljno malom intervalu oko  $\lambda_{(k)}$ . Neka je  $s > 1$ . Definiramo skup

$$\Omega_n = \Omega_n(k, s) = \{a_{np}^{-2} |\lambda_{(k)} - \lambda_{(i)}| > s\varepsilon^{(n)} : i \neq k = 1, 2, \dots, p\}.$$

Iz (2.11) slijedi  $s\varepsilon^{(n)} \rightarrow 0$ . Sada, koristeći metodu koju su Heiny i Mikosch iskoristili u [13], zaključujemo

$$\lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n^c) = \lim_{n \rightarrow \infty} \mathbb{P}(a_{np}^{-2} \min\{\lambda_{(k-1)} - \lambda_{(k)}, \lambda_{(k)} - \lambda_{(k+1)}\} \leq s\varepsilon^{(n)}) = 0.$$

Po propoziciji 2.2.2 jedinični svojstveni vektor  $\mathbf{v}_k$  pridružen svojstvenoj vrijednosti  $\lambda_{(k)}$  i projicirani vektor  $\mathbf{P}_{\mathbf{e}_{L_k}}(\mathbf{v}_k) = (\mathbf{v}_k)_{L_k} \mathbf{e}_{L_k}$  zadovoljavaju za fiksni  $\delta > 0$  sljedeće:

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(\|\mathbf{v}_k - (\mathbf{v}_k)_{L_k} \mathbf{e}_{L_k}\|_{\ell_2} > \delta) \\ \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\{\|\mathbf{v}_k - (\mathbf{v}_k)_{L_k} \mathbf{e}_{L_k}\|_{\ell_2} > \delta\} \cap \Omega_n) + \limsup_{n \rightarrow \infty} \mathbb{P}(\Omega_n^c) \\ \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\{2\varepsilon^{(n)} / (s\varepsilon^{(n)} - \varepsilon^{(n)}) > \delta\} \cap \Omega_n) \\ \leq \limsup_{n \rightarrow \infty} \mathbb{P}(\{2/(s-1) > \delta\}) = \mathbf{1}_{\{2/(s-1) > \delta\}}. \end{aligned} \quad (2.12)$$

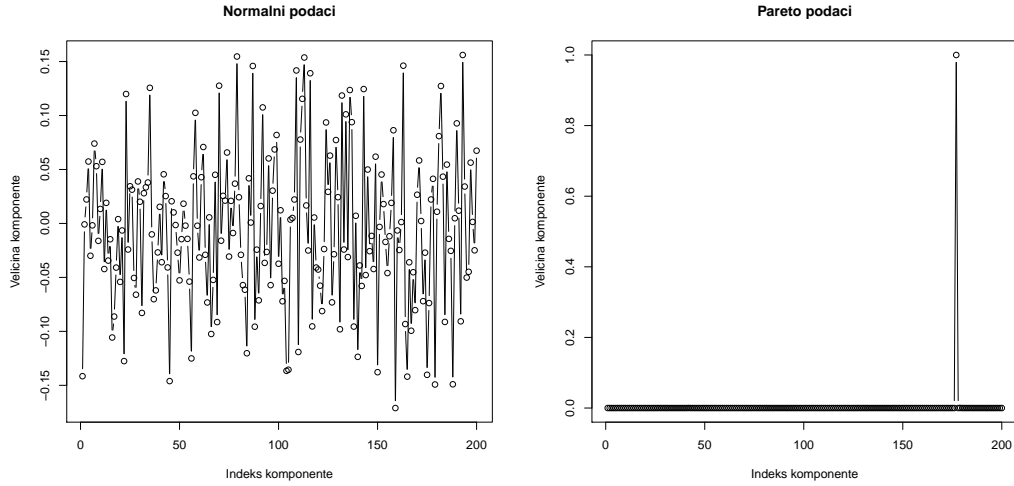
Desna strana u (2.12) je 0 za dovoljno velik  $s$ . S obzirom da su i  $\mathbf{v}_k$  i  $\mathbf{e}_{L_k}$  jedinični vektori slijedi

$$\|\mathbf{v}_k - \mathbf{e}_{L_k}\|_{\ell_2} \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty,$$

što dokazuje tvrdnju za svojstvene vektore.  $\square$

Kao što smo vidjeli, postoji bliska veza između vektora  $\mathbf{v}_j$  i  $\mathbf{e}_{L_j}$ . Ti vektori dijele još jedno svojstvo, oni su lokalizirani. Za vektor kažemo da je *lokaliziran* ako je koncentriran u samo nekoliko komponenti. U suprotnom, ako nije lokaliziran, kažemo da je *delokaliziran*. Na sljedećim grafovima ilustrirat ćemo lokaliziranost. Na slici 2.2 se nalaze komponente svojstvenog vektora pridruženog najvećoj svojstvenoj vrijednosti matrice  $\mathbf{Z}\mathbf{Z}^t$  za simulirane matrice podataka  $\mathbf{Z}$ . Na desnom grafu se nalaze komponente svojstvenog vektora koji je dobiven za matricu podataka čije su komponente nezavisno jednako distribuirane iz Pareto(0.8) razdiobe. Na grafu uočavamo da je samo jedna od  $p = 200$  komponenti značajno različita od 0, pa možemo zaključiti da je lokaliziran. Na lijevom grafu nalazi se svojstveni vektor za najveću svojstvenu vrijednost kada su komponente također nezavisne i jednako distribuirane standardne normalne. Za razliku od slučaja kada su komponente matrice iz Pareto razdiobe, na lijevom grafu vidimo da je mnogo komponenti

svojstvenog vektora slične veličine, stoga zaključujemo da je delokaliziran. Pokazuje se da kada komponente matrice podataka dolaze iz distribucije s beskonačnim četvrtim momentom, svojstveni vektori su obično lokalizirani. O slučaju Wignerove matrice više se može pronaći u [6].



Slika 2.2: Komponente vektora  $\mathbf{v}_1$ . Na lijevom grafu se nalaze podaci u slučaju kada su komponente matrice podataka iz standardne normalne distribucije, dok se na desnom nalaze u slučaju kada su iz Pareto(0.8) distribucije. Veličina matrice podataka za oba grafa je  $p \times n$  uz  $p = 200$  i  $n = 1000$ .

### 2.3 Autokovarijacijske matrice

Zbog važnosti teme u analizi višedimenzionalnih vremenskih nizova i široke primjene, u ovom odjeljku proširujemo dosadašnje rezultate na autokovarijacijske matrice.

Neka je  $(Z_{it})$  polje podataka,  $p \times n$  matricu konstruiramo na sljedeći način:

$$\mathbf{Z}(s, k) = \mathbf{Z}_n(s, k) = (Z_{i-s, t-k})_{i=1,2,\dots,p; t=1,2,\dots,n}, \quad s, k \in \mathbb{Z}.$$

Generalizirana uzoračka autokovarijacijska matrica (nenormalizirana) definirana je sa

$$(\mathbf{Z}(0, 0)\mathbf{Z}(s, k)^t), \quad s, k \in \mathbb{Z},$$

čije su komponente

$$(\mathbf{Z}(0, 0)\mathbf{Z}(s, k)^t)_{ij} = \sum_{t=1}^n Z_{i,t} Z_{j-s, t-k}, \quad i, j = 1, 2, \dots, p.$$

Uočimo, ako je  $\min(|s|, |k|) \neq 0$ , generalizirana autokovarijacijska matrica neće biti simetrična i može sadržavati kompleksne svojstvene vrijednosti u svom spektru. Za daljnju analizu koristit ćemo *singularne vrijednosti*. *Singularna vrijednost* kvadratne matrice  $\mathbf{Z}$  definira se kao korijen svojstvene vrijednosti matrice  $\mathbf{Z}\mathbf{Z}^t$ . U skladu s prijašnjom notacijom singularne vrijednosti označavat ćemo s  $(\lambda_i(s, k))$ , a pridružene uređajne statistike s  $\lambda_{(1)}(s, k) \geq \lambda_{(2)}(s, k) \geq \dots \geq \lambda_{(p)}(s, k)$ .

**Teorem 2.3.1.** *Neka su  $s, k \in \mathbb{Z}$ , a  $\mathbf{Z}(0, 0)$  i  $\mathbf{Z}(s, k)$   $p \times n$  matrice s nezavisno jednako distribuiranim komponentama. Pretpostavljamo da su zadovoljeni sljedeći uvjeti:*

- *Uvjet regularne varijacije (1.5) za neki  $\alpha \in (0, 4)$ .*
- *$\mathbb{E}[Z] = 0$  za  $\alpha \geq 2$ .*
- *Niz prirodnih brojeva  $(p_n)$  raste kao u (2.1) za neki  $\beta \geq 0$ .*

1. *Ako je  $k \neq 0$ , tada*

$$a_{np}^{-2} \lambda_{(1)}(s, k) \xrightarrow{\mathbb{P}} 0.$$

*Pretpostavimo sada da je  $k = 0$ , tada vrijede sljedeći rezultati:*

2. *Ako je  $\beta \in [0, 1]$ , tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,p-|s|} |\lambda_{(i)}(s, 0) - D_{(i)}^{\rightarrow}| \xrightarrow{\mathbb{P}} 0. \quad (2.13)$$

3. *Ako je  $\beta > 1$ , tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,n-|s|} |\lambda_{(i)}(s, 0) - D_{(i)}^{\downarrow}| \xrightarrow{\mathbb{P}} 0. \quad (2.14)$$

4. *Ako je  $\min(\beta, \beta^{-1}) \in ((\alpha/2 - 1)_+, 1]$ , tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,p-|s|} |\lambda_{(i)}(s, 0) - Z_{(i),np}^2| \xrightarrow{\mathbb{P}} 0. \quad (2.15)$$

*Dokaz.* U dokazu promatramo slučaj kada je  $\beta \in [0, 1]$ , jer slučaj kada je  $\beta > 1$  slijedi kao i prije prelaskom sa  $\mathbf{Z}\mathbf{Z}^t$  na  $\mathbf{Z}^t\mathbf{Z}$  korištenjem Weylove nejednakosti (1.16). Ideja dokaza je svesti  $\mathbf{Z}\mathbf{Z}^t$  na dijagonalnu matricu. U slučaju kada je  $k = 0$ , matricu  $\mathbf{Z}(0, 0)\mathbf{Z}(s, k)^t$  reduciramo na  $p \times p$  matricu  $\mathbf{M}^{(s,k)}$  koja ima vrijednosti različite od 0 samo na  $s$ -toj sporednoj dijagonali. Komponente  $s$ -te sporedne dijagonale matrice  $\mathbf{M}^{(s,k)}$  su  $\mathbf{M}_{i,i+s}^{(s,k)}$ ,  $i = 1 + s_-, \dots, p - s_+$ , gdje su  $s_+, s_- \geq 0$  pozitivan i negativan dio od  $s$ .

Neka je  $k \in \mathbb{Z}$ . Zbog jednostavnosti zapisa pretpostavljamo još i da je  $s \geq 0$ . Matricu  $\mathbf{M}^{(s,k)}$  definiramo na sljedeći način:

$$\mathbf{M}_{i,i+s}^{(s,k)} = \mathbf{1}_{\{k=0\}} (\mathbf{Z}(0,0)\mathbf{Z}(s,k)^t)_{i,i+s} = \mathbf{1}_{\{k=0\}} \sum_{t=1}^n Z_{it}^2, \quad i = 1, 2, \dots, p-s,$$

i  $\mathbf{M}_{ij}^{(s,k)} = 0$  za sve ostale  $i, j$ . Sada imamo

$$\begin{aligned} & ((\mathbf{Z}(0,0)\mathbf{Z}(s,k)^t - \mathbf{M}^{(s,k)})(\mathbf{Z}(0,0)\mathbf{Z}(s,k)^t - \mathbf{M}^{(s,k)})^t)_{ij} \\ &= \sum_{u=1}^p \sum_{t_1=1}^n \sum_{t_2=1}^n Z_{i,t_1} Z_{j,t_2} Z_{u-s,t_1-k} Z_{u-s,t_2-k} \mathbf{1}_{\{i \neq u-s, j \neq u-s\}} \\ & \quad \times (\mathbf{1}_{\{i=j\}} + \mathbf{1}_{\{i \neq j, t_1=t_2\}} + \mathbf{1}_{\{i \neq j, t_1 \neq t_2\}}) \\ &= \mathbf{D}_{ij} + \mathbf{F}_{ij} + \mathbf{R}_{ij}. \end{aligned} \tag{2.16}$$

Po lemi 4.1 u [13] za dekompoziciju u (2.16) vrijedi

$$a_{np}^{-4} \|\mathbf{D} + \mathbf{F} + \mathbf{R}\|_2^2 \xrightarrow{\mathbb{P}} 0.$$

Iz toga slijedi

$$\begin{aligned} & a_{np}^{-4} \|\mathbf{Z}(0,0)\mathbf{Z}(s,k)^t - \mathbf{M}^{(s,k)}\|_2^2 \\ &= a_{np}^{-4} \|(\mathbf{Z}(0,0)\mathbf{Z}(s,k)^t - \mathbf{M}^{(s,k)})(\mathbf{Z}(0,0)\mathbf{Z}(s,k)^t - \mathbf{M}^{(s,k)})^t\|_2 \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Time je dovršen dokaz prve tvrdnje. Kako s vjerojatnošću koja teži k 1, matrica  $\mathbf{M}^{(s,k)}$  ima tražene singularne vrijednosti, druga tvrdnja slijedi iz prve preko Weylove nejednakosti (1.16). Dokaz četvrte tvrdnje se može pronaći u [13].  $\square$

## Poglavlje 3

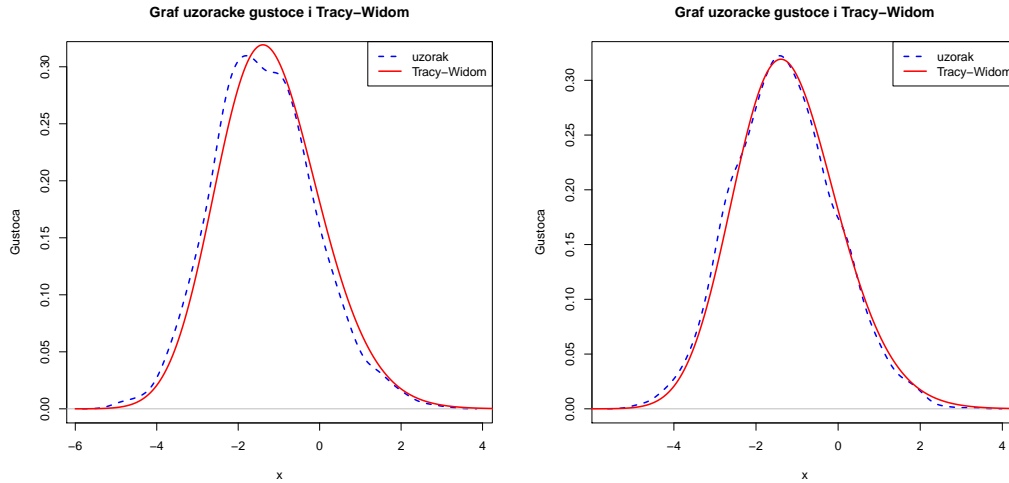
# Uzoračke autokovarijacijske matrice

U ovom poglavlju bit će prikazani neki rezultati o ekstremnim vrijednostima uzoračke kovarijacijske i autokovarijacijske matrice višedimenzionalnih vremenskih nizova teškog repa. Zavisnost u strukturi višedimenzionalnih podataka često se proučava kroz kovarijacijsku matricu koja je tipično procijenjena svojim uzoračkim analogonom. Na primjer, u analizi glavnih komponenti, glavne komponente čine upravo svojstveni vektori uzoračke kovarijacijske matrice koji su pridruženi najvećim svojstvenim vrijednostima, dok veličina svojstvenih vrijednosti daje empirijsku mjeru važnosti tih komponenti.

Također, u ovom poglavlju teorijski rezultati bit će ilustrirani na simuliranim i empirijskim podacima. Poglavlje započinjemo ilustracijama rezultata iz prethodnih poglavlja.

### Slučaj distribucije lakog repa

Kao što je prikazano u prethodnim poglavljima analiza višedimenzionalnih vremenskih nizova s nezavisno jednako distribuiranim komponentama je puno razvijenija u slučaju kada se radi o distribuciji lakog repa naspram one kod slučaja distribucije teškog repa. U prvom poglavlju vidjeli smo kako je Johnstone u [16] dokazao konvergenciju po distribuciji najveće svojstvene vrijednosti uzoračke kovarijacijske matrice uz pogodnu normalizaciju prema *Tracy-Widom distribuciji*. U slučaju gaussovske matrice podataka treba uzeti u obzir da je račun uvelike olakšan činjenicom da je matrica invarijantna na ortogonalne transformacije. Kod matrica čiji podaci nisu gaussovski i matrica nije invarijantna na ortogonalne transformacije, razvoj teorije je išao puno sporije zbog komplikacija u računu. U radu [24] koji su napisali Tao i Vu, rezultat je proširen na matricu  $\mathbf{Z}$  čije su komponente nezavisno jednako distribuirane ali ne i normalne. Nadalje, oni su u svom radu pokazali konvergenciju tehnikom u kojoj korak po korak komponente koje su normalne i za koje vrijedi konvergencija po distribuciji zamjenjuju komponentama koje nisu normalne, ali imaju ista prva četiri momenta kao odgovarajuća normalna slučajna varijabla. Njihov rezultat je posljedica takozvanog *Teorema o četiri momenta* (eng. *Four Moment Theorem*) koji opisuje

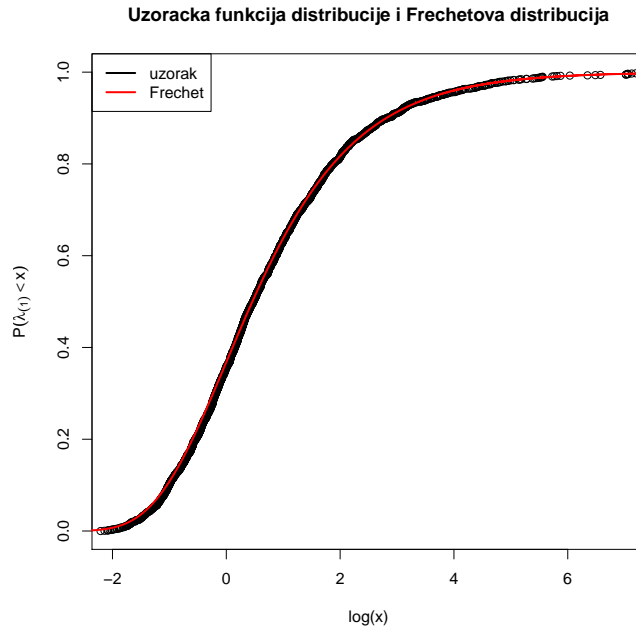


(a) Komponente matrice  $\mathbf{Z}$  iz standardne nor- (b) Distribucija komponenti matrice  $\mathbf{Z}$  na  
mane razdiobe. grafu je  $\mathbb{P}(\mathbf{Z} = -\sqrt{3}) = \mathbb{P}(\mathbf{Z} = \sqrt{3}) = 1/6$ ,  $\mathbb{P}(\mathbf{Z} = 0) = 2/3$ .

Slika 3.1: Na grafovima su prikazane procjenjene uzoračke funkcije gustoće najveće svojstvene vrijednosti, te su uspoređene s funkcijom gustoće Tracy-Widom distribucije. Uzorak je dobiven na temelju 2000 simulacija matrice podataka  $\mathbf{Z}$  dimenzije  $200 \times 1000$ .

neosjetljivost svojstvenih vrijednosti na promjenu distribucije komponenti. Za razliku od prethodno navedenog, u slučaju kada se radi o razdiobi koja nema konačan četvrti moment, normalizirana najveća svojstvena vrijednost prati *Fréchetovu distribuciju* koja je definirana u (1.21). Na slici 3.1 uspoređene su uzoračke funkcije gustoće u slučaju kada su komponente iz standardne normalne distribucije i kada nisu. Uočimo da su prva četiri momenta ( $\mathbb{E}[\mathbf{Z}] = 0$ ,  $\mathbb{E}[\mathbf{Z}^2] = 1$ ,  $\mathbb{E}[\mathbf{Z}^3] = 0$ ,  $\mathbb{E}[\mathbf{Z}^4] = 3$ ) distribucije čija je uzoračka gustoća prikazana na desnoj slici upravo jednaka onima standardne normalne. U oba slučaja, na grafovima se vidi da aproksimacija *Tracy-Widom* distribucijom vrlo dobro funkcionira. Na slici 3.2 ilustrirano je kako ovaj pristup nije dobar u slučaju kada ne postoji četvrti moment, tj.  $\mathbb{E}[\mathbf{Z}^4] = \infty$ .

Na slici 3.1 uspoređena je uzoračka funkcija gustoće normaliziranih najvećih svojstvenih vrijednosti koja je dobivena na temelju 2000 simulacija uzoračke kovarijacijske matrice  $\mathbf{Z}\mathbf{Z}^t$  ( $n = 1000$ ,  $p = 200$ ) s Tracy-Widom gustoćom. Ukoliko slučajna varijabla  $Z$  nema konačan četvrti moment i ako je zadovoljen uvjet regularne varijacije (1.5) onda za graničnu distribuciju, umjesto Tracy-Widom, uzimamo *Fréchetovu distribuciju* po (1.20). Slika 3.2 ilustrira ovu činjenicu na temelju simulacije uzoračke kovarijacijske matrice  $\mathbf{Z}\mathbf{Z}^t$  tako da komponente matrice podataka  $\mathbf{Z}$  dolaze iz distribucije (2.10).



Slika 3.2: Na grafu su uspoređene uzoracka funkcija distribucije najveće svojstvene vrijednosti  $\lambda_{(1)}$  i Fréchetova distribucija za  $\alpha = 1.6$ . Rezultat se temelji na 2000 simulacija matrice dimenzije  $200 \times 1000$  s nezavisno jednako distribuiranim komponentama koje nemaju konačan četvrti moment.

### Slučaj teškog repa

U ovom odjeljku fokus prebacujemo s distribucija koje imaju konačan četvrti moment na one kod kojih je on nije konačan. Pokazuje se da pretpostavka o konačnom četvrtom momentu može biti narušena kada se analiziraju podaci povezani s financijama, telekomunikacijama i osiguranjima. Na slici 3.3 prikazani su donji i gornji repni indeks ( $\alpha_L, \alpha_U$ ) za  $p = 470$  vremenskih nizova log-povrata komponenti od S&P 500 indeksa procijenjenih na temelju  $n = 1259$  dnevnih opažanja iz perioda od 8. veljače 2013. godine do 7. veljače 2018. godine. Rezultat je dobiven uz pretpostavku da se za svaki red  $(Z_{it})_{t=1,\dots,n}$  od  $\mathbf{Z}$  rep ponaša kao

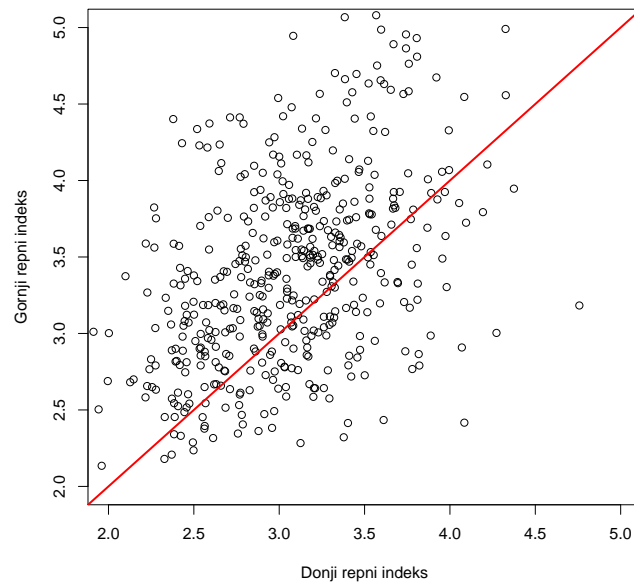
$$\mathbb{P}(Z_{it} > x) \sim c_U x^{-\alpha_U} \quad i \quad \mathbb{P}(Z_{it} < -x) \sim c_L x^{-\alpha_L}, \quad x \rightarrow \infty,$$

za neke nenegativne konstante  $c_U$  i  $c_L$ . Repne indekse dobitaka i gubitaka procjenjujemo tzv. *Hillovim procjeniteljem*. Za uređajne statistike  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  koristeći  $k$  najvećih



vrijednosti u uzorku *Hillov procjenitelj* za repni indeks  $\alpha$  je definiran formulom

$$\hat{\alpha}_{n,k} = \left( \frac{1}{k} \sum_{j=1}^k \log \frac{X_{(n-j+1)}}{X_{(n-k+1)}} \right)^{-1}.$$



Slika 3.3: Repni indeksi log-povrata od 470 vremenskih nizova koji čine S&P 500 index. Procijenjene vrijednosti  $(\hat{\alpha}_L, \hat{\alpha}_U)$  donjeg i gornjeg repnog indeksa  $(\alpha_L, \alpha_U)$  dobivene su tzv. *Hillovim procjeniteljem*. Na grafu je naznačen još i pravac  $\hat{\alpha}_L = \hat{\alpha}_U$ .

Treba napomenuti da je procjena relativno gruba jer nisu uzeti u obzir ni zavisnosti ni nestacionarnosti podataka. Na grafu se jasno vidi da većina pozitivnih kao i negativnih povrata, tj. gubitaka ima repne indekse manje od 4 što odgovara nepostojanju četvrtog momenta. Ponašanje najveće svojstvene vrijednosti značajno se mijenja kada  $\mathbf{Z}$  nema konačan četvrti moment.

### 3.1 Konvergencija točkovnih procesa

Kao što je ranije spomenuto, Soshnikov [22, 23] je prvi počeo proučavati ponašanje najveće svojstvene vrijednosti  $\lambda_{(1)}$  u slučaju regularno varirajućih komponenti  $Z_{it}$  za indeks  $\alpha \in$

(0, 2). On je pokazao konvergenciju točkovnog procesa (u smislu konvergencije po distribucije točkovnih procesa iz odjeljka 1.6)

$$N_n = \sum_{i=1}^p \varepsilon_{a_{np}^{-2}\lambda_i} \xrightarrow{d} N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}, \quad n \rightarrow \infty, \quad (3.1)$$

pod uvjetom (1.2) za  $(p_n)$ . Ovdje je

$$\Gamma_i = E_1 + E_2 + \cdots + E_i, \quad i \geq 1, \quad (3.2)$$

gdje je  $(E_i)$  niz nezavisno jednako distribuiranih eksponencijalnih slučajnih varijabli, a  $\varepsilon_y$  je Diracova mjera u  $y$ . Drugim riječima,  $N$  je Poissonov točkovni proces na  $(0, \infty)$  s mjerom očekivanja  $\mu(x, \infty) = x^{-\alpha/2}$ ,  $x > 0$ . Iz (3.1) je sada lako doći do distribucije od  $a_{np}^{-2}\lambda_{(k)}$  za fiksni  $k \geq 1$ :

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(a_{np}^{-2}\lambda_{(k)} \leq x) &= \lim_{n \rightarrow \infty} \mathbb{P}(N_n(x, \infty) < k) = \mathbb{P}(N(x, \infty) < k) = \mathbb{P}(\Gamma_k^{-2/\alpha} \leq x) \\ &= \sum_{s=1}^{k-1} \frac{(\mu(x, \infty))^s}{s!} e^{-\mu(x, \infty)}, \quad x > 0. \end{aligned}$$

Sada, za  $k = 1$ , tj. za najveću svojstvenu vrijednost slijedi

$$\frac{\lambda_{(1)}}{a_{np}^2} \xrightarrow{d} \Gamma_1^{-\alpha/2}, \quad n \rightarrow \infty,$$

gdje granična vrijednost ima *Fréchetovu distribuciju* s parametrom  $\alpha/2$  i pripadnu funkciju distribucije

$$\Phi_{\alpha/2}(x) = e^{-x^{-\alpha/2}}, \quad x > 0. \quad (3.3)$$

Napomenimo još da se uvjet regularne varijacije (1.5) u ovom slučaju može zamijeniti slabijom pretpostavkom  $\mathbb{P}(|Z| > x) = L(x)x^{-\alpha}$  za neku sporo varirajuću funkciju  $L$  koja zadovoljava (1.4). Zajedničku konvergenciju viših uređajnih statistika dobivamo kao posljedicu konvergencije iz (3.1) i teorema o neprekidnom preslikavanju: za svaki  $k \geq 1$ ,

$$a_{np}^{-2}(\lambda_{(1)}, \lambda_{(2)}, \dots, \lambda_{(k)}) \xrightarrow{d} (\Gamma_1^{-\alpha/2}, \Gamma_2^{-\alpha/2}, \dots, \Gamma_k^{-\alpha/2}), \quad n \rightarrow \infty.$$

Iz teorije o točkovnim procesima za nezavisno jednako distribuirane točke (vidi Resnick [21, 20]) slijedi da (3.1) vrijedi i u slučaju ako  $N_n$  zamijenimo s točkovnim procesom  $\sum_{i=1}^p \sum_{t=1}^n \varepsilon_{a_{np}^{-2}Z_{it}^2}$ . Tada također imamo za svaki  $k \geq 1$ ,

$$a_{np}^{-2}(Z_{(1),np}^2, Z_{(2),np}^2, \dots, Z_{(k),np}^2) \xrightarrow{d} (\Gamma_1^{-\alpha/2}, \Gamma_2^{-\alpha/2}, \dots, \Gamma_k^{-\alpha/2}), \quad n \rightarrow \infty. \quad (3.4)$$

Kao što je spomenuto ranije, ovaj rezultat je kasnije proširen i na  $\alpha \in (2, 4)$ , kao i na slučaj kada su retci matrice  $\mathbf{Z}$  nezavisno jednako distribuirani linearni procesi s regularno varirajućim šumom.

U nastavku dajemo ilustracije kako se aproksimacije iz teorema 2.1.1 mogu iskoristiti za asimptotsku konvergenciju najveće svojstvene vrijednosti matrice  $\mathbf{ZZ}'$  preko konvergencije odgovarajućeg točkovnog procesa.

**Lema 3.1.1.** *Neka je  $\mathbf{Z}$  matrica dimenzije  $p \times n$  s nezavisno jednako distribuiranim komponentama. Pretpostavljamo da su zadovoljeni sljedeći uvjeti:*

- *Uvjet regularne varijacije (1.5) za neki  $\alpha \in (0, 4)$ .*
- *$\mathbb{E}[Z] = 0$  za  $\alpha \geq 2$ .*
- *Niz prirodnih brojeva  $(p_n)$  raste kao u (2.1) za neki  $\beta \geq 0$ .*

1. *Ako je  $\beta \geq 0$ , tada*

$$\sum_{i=1}^p \varepsilon_{a_{np}^{-2}(D_i^{\rightarrow} - c_n)} \xrightarrow{d} N, \quad n \rightarrow \infty, \quad (3.5)$$

*gdje  $c_n = 0$  ako je  $\mathbb{E}[D^{\rightarrow}] = 0$  i  $c_n = \mathbb{E}[D^{\rightarrow}] = n\mathbb{E}[Z^2]$  u suprotnom.*

2. *Ako je  $\beta \geq 0$ , tada*

$$\sum_{i=1}^p \varepsilon_{a_{np}^{-2}(Z_{(i),np}^{\rightarrow})} \xrightarrow{d} N, \quad n \rightarrow \infty, \quad (3.6)$$

*Konvergencija točkovnih procesa vrijedi na prostoru točkovnih mjera s prostorom stanja  $(0, \infty)$  kojemu je pridružena vague topologija (vidi odjeljak 1.6 za koncept konvergencije).*

**Napomena 3.1.2.** *Dokaz teorema može se pronaći radu [13] Heinyja i Mikoscha. Slični rezultati su također korišteni u radovima [9, 10] Davisa i drugih. Nadalje, centriranje  $c_n$  u slučaju konačne varijance se može izbjeći ako  $n/a_{np}^2 \rightarrow 0$ . Taj uvjet je zadovoljen ako je  $\beta > \alpha/2 - 1$ .*

Pomoću teorema 2.1.1 i argumenata koje su koristili Davis i njegovi kolege u radovima [9, 10] dolazimo do konvergencije točkovnih procesa normaliziranih svojstvenih vrijednosti.

**Teorem 3.1.3.** *Pretpostavimo da vrijede uvjeti leme 3.1.1. Ako je zadovoljen uvjet*

$$\min(\beta, \beta^{-1}) \in ((\alpha/2 - 1)_+, 1] \quad (3.7)$$

tada

$$\sum_{i=1}^p \varepsilon_{a_{np}^{-2}\lambda_i} \xrightarrow{d} N. \quad (3.8)$$

u prostoru točkovnih mjera s prostorom stanja  $(0, \infty)$  kojemu je pridružena vague topologija (vidi odjeljak 1.6 za detalje i koncept konvergencije).

*Dokaz.* Granična relacija (3.8) slijedi iz (3.6) u kombinaciji s (2.9). Alternativno, možemo iskoristiti relaciju (3.5) za oba niza  $(D_i^{\rightarrow})$  i  $(D_i^{\downarrow})$  (uočimo, konvergencija drugog niza je dobivena na sličan način kao i prije, zamjenom uloga  $n$  i  $p$ ), činjenicu da  $(\max(n, p)/a_{np}^2 \rightarrow 0$  ako je  $\min(\beta, \beta^{-1}) \in ((\alpha/2 - 1)_+, 1]$  (stoga se centriranje točaka  $(D_i^{\rightarrow})$  i  $(D_i^{\downarrow})$  u (3.5) može izbjeći zbog  $\mathbb{E}[Z^2] < \infty$ ) i aproksimacije (2.7) ili (2.8).  $\square$

Kao posljedicu konvergencije normaliziranih svojstvenih vrijednosti matrice  $\mathbf{ZZ}^t$  prethodnog teorema Heiny i Mikosch u [13] dobivaju sljedeće rezultate. Direktna posljedica teorema, za svaki fiksni  $k \geq 1$ , je

$$a_{np}^{-2}(\lambda_{(1)}, \lambda_{(2)}, \dots, \lambda_{(k)}) \xrightarrow{d} (\Gamma_1^{-\alpha/2}, \Gamma_2^{-\alpha/2}, \dots, \Gamma_k^{-\alpha/2}) \quad (3.9)$$

Koristeći metode koje su zajedno s Davisom koristili u [9], Heiny i Mikosch u [13] dokazuju za  $\alpha \in (2, 4)$

$$a_{np}^{-2}(\lambda_{(1)} - (\max(p, n))\mathbb{E}[Z^2], \dots, \lambda_{(k)} - (\max(p, n))\mathbb{E}[Z^2]) \xrightarrow{d} (\Gamma_1^{-\alpha/2}, \Gamma_2^{-\alpha/2}, \dots, \Gamma_k^{-\alpha/2}). \quad (3.10)$$

Relacije (3.9) i (3.10) povlače da za  $\alpha \in (0, 4)$  i za svaki fiksni  $k \geq 1$  vrijedi

$$a_{np}^{-2}(\lambda_{(1)} - \lambda_{(2)}, \dots, \lambda_{(k)} - \lambda_{(k+1)}) \xrightarrow{d} (\Gamma_1^{-\alpha/2} - \Gamma_2^{-\alpha/2}, \dots, \Gamma_k^{-\alpha/2} - \Gamma_{k+1}^{-\alpha/2}). \quad (3.11)$$

Povezani rezultati mogu se dobiti i za rastući broj uređajnih statistika. Na primjer, zajednička konvergencija najveće svojstvene vrijednosti  $a_{np}^{-2}\lambda_{(1)}$  i traga  $a_{np}^{-2}(\lambda_1 + \lambda_2 + \dots + \lambda_p)$ . Preciznije, za  $\alpha \in (0, 2)$  pod uvjetima leme 3.1.1 vrijedi

$$\frac{\lambda_{(1)}}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \xrightarrow{d} \frac{\Gamma_1^{-\alpha/2}}{\Gamma_1^{-\alpha/2} + \Gamma_2^{-\alpha/2} + \dots + \Gamma_p^{-\alpha/2}}.$$

Više primjera i detaljni dokazi mogu se pronaći u radu [10] Davisa i njegovih kolega.

## 3.2 Zavisnost između redaka i stupaca

Detaljna matematička analiza i rezultati iz ovog odjeljka mogu se pronaći u radovima Davisa i njegovih kolega [9] i [10], radu Heinyja i Mikoscha [13] te radu [14] Heinyja, Mikoscha i Davisa.

### Model

Kod proučavanja kovarijacijskih matrica višedimenzionalnih vremenskih nizova  $(\mathbf{Z}_n)$ , prirodno je pretpostaviti da postoji zavisnost između komponenti  $(Z_{it})$ . Model koji dopušta *linearnu zavisnost* između redaka i stupaca od  $\mathbf{Z}$  definiran je relacijom

$$Z_{it} = \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} h_{kl} X_{i-k, t-l}, \quad i, t \in \mathbb{Z}, \quad (3.12)$$

gdje je  $(X_{it})_{i,t \in \mathbb{Z}}$  polje nezavisno jednako distribuiranih slučajnih varijabli a  $(h_{kl})_{k,l \in \mathbb{Z}}$  je polje realnih brojeva. Bez obzira na to što je linearna zavisnost restriktivna pretpostavka, ta zavisnost omogućava otkrivanje dijelova kovarijacijske matrice koji su odrednice najvećih svojstvenih vrijednosti. Ako nizovi u (3.12) konvergiraju g.s., tada je  $(Z_{it})$  jako stacionarno slučajno polje. Reprezentante polja  $\mathbf{X}$  i  $\mathbf{Z}$  označavat ćemo s  $X$  i  $Z$ . Pretpostavimo da je  $X$  regularno varirajuća, tj. zadovoljava uvjet (1.5). Pretpostavimo također da vrijedi  $\mathbb{E}[X] = 0$  kad god  $\mathbb{E}[X^2] < \infty$ . Nadalje, da bi osigurali g.s. apsolutnu konvergenciju nizova u (3.12) potreban nam je uvjet sumabilnosti

$$\sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |h_{kl}|^\delta < \infty \quad (3.13)$$

za neki  $\delta \in (0, \min(\alpha/2, 1))$ . Uz uvjete regularne varijacije na  $X$  i sumabilnosti (3.13), granične i konačnodimenzionalne distribucije polja  $(Z_{it})$  su regularno varirajuće s indeksom  $\alpha$ , stoga kažemo da su polja  $(Z_{it})$  i  $(X_{it})$  regularno varirajuća.

Model (3.12) je predstavio Davis sa svojim kolegama u radu [11], pretpostavljajući da su redovi nezavisno jednako distribuirani, te u sadašnjoj formi u svom radu [10].

### Uzoračka kovarijacijska i autokovarijacijska matrica

Iz polja  $(Z_{it})$  konstruiramo matrice dimenzije  $p \times n$  relacijom

$$\mathbf{Z}_n(s) = (Z_{i,t+s})_{i=1,2,\dots,p; t=1,2,\dots,n}, \quad s = 0, 1, 2, \dots \quad (3.14)$$

Kao i prije pisat ćemo  $\mathbf{Z} = \mathbf{Z}_n(0)$ . Nenormalizirane *uzoračke autokovarijacijske matrice* dana su sljedećom relacijom

$$\mathbf{Z}_n(0)\mathbf{Z}_n(s)^t, \quad s = 0, 1, 2, \dots \quad (3.15)$$

$s$  iz prethodne relacije nazivamo vremenski zaostatak (eng. *lag*). Za  $s = 0$  dobivamo *uzoračku kovarijacijsku matricu*.

U nastavku ćemo promatrati asimptotska svojstva (funkcija) svojstvenih i singularnih vrijednosti kovarijacijskih i autokovarijacijskih matrica. Prisjetimo se, singularne vrijednosti matrice  $A$  su korijeni svojstvenih vrijednosti pozitivno semidefinitne matrice  $AA^t$  i njezina spektralna norma  $\|A\|_2$  je upravo jednaka najvećoj singularnoj vrijednosti. Primijetimo,  $\mathbf{Z}_n(0)\mathbf{Z}_n(s)^t$  nije simetrična pa stoga njezine svojstvene vrijednosti mogu biti kompleksne. Da bi pojednostavili račun, promatrat ćemo svojstvene vrijednosti matrice  $\mathbf{Z}_n(0)\mathbf{Z}_n(s)^t\mathbf{Z}_n(s)\mathbf{Z}_n(0)^t$ , koje su kvadrati singularnih vrijednosti matrice  $\mathbf{Z}_n(0)\mathbf{Z}_n(s)^t$ . Nadalje, za fiksne cijele brojeve  $n \geq 1$  i  $s \geq 0$  reciklirat ćemo  $\lambda$ -notaciju za singularne vrijednosti  $\lambda_1(s), \lambda_2(s), \dots, \lambda_p(s)$  autokovarijacijske matrice  $\mathbf{Z}_n(0)\mathbf{Z}_n(s)^t$ , radi suzbijanja ovisnosti o  $n$ . Pripadne uređajne statistike označavat ćemo s

$$\lambda_{(1)}(s), \lambda_{(2)}(s), \dots, \lambda_{(p)}(s). \quad (3.16)$$

U slučaju kada je  $s = 0$  koristit ćemo oznaku  $\lambda_i$  umjesto  $\lambda_i(0)$ .

## Notacija i aproksimacije

Definiramo pomoćnu matricu izvedenu iz matrice koeficijenata  $\mathbf{H} = (h_{kl})_{k,l \in \mathbb{Z}}$  s:

$$\mathbf{H}(s) = (h_{k,l+s})_{k,l \in \mathbb{Z}}, \quad \mathbf{M}(s) = \mathbf{H}(0)\mathbf{H}(s)^t, \quad s \geq 0.$$

Uočimo

$$(\mathbf{M}(s))_{ij} = \sum_{l \in \mathbb{Z}} h_{i,l} h_{j,l+s}, \quad i, j \in \mathbb{Z}. \quad (3.17)$$

Singularne vrijednosti u padajućem poretku matrice  $\mathbf{M}(s)$  označimo s:

$$v_1(s) \geq v_2(s) \geq \dots \quad (3.18)$$

Neka je  $r(s)$  rang matrice  $\mathbf{M}(s)$ . Tada je  $v_{r(s)} > 0$  i  $v_{r(s)+1} = 0$ , ako je  $r(s) < \infty$ . U suprotnom  $v_i(s) > 0$  za svaki  $i$ . Također, u skladu s prethodnom notacijom, pišemo  $r = r(0)$ .

Uz uvjet sumabilnosti (3.13) za  $(h_{kl})$  i fiksni  $s \geq 0$ ,

$$\begin{aligned} \sum_{i=1}^{\infty} (v_i(s))^2 &= \|\mathbf{M}(s)\|_F^2 = \sum_{i,j \in \mathbb{Z}} \sum_{l_1, l_2 \in \mathbb{Z}} h_{i,l_1} h_{j,l_1+s} h_{i,l_2} h_{j,l_2+s} \\ &\leq c \left( \sum_{l_1, l_2 \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} |h_{i,l_1} h_{i,l_2}| \right)^2 \leq c \sum_{l_1 \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} |h_{i,l_1}| < \infty. \end{aligned} \quad (3.19)$$

Stoga, sve singularne vrijednosti su konačne pa je uređaj u (3.18) opravdan.

**Napomena 3.2.1.** U (3.19) i u nastavku koristimo oznaku  $c$  za konstantu čiji iznos nije relevantan za rezultate.

Singularne vrijednosti  $\lambda_i(s)$  aproksimirat ćemo u terminima  $p$  najvećih vrijednosti, u padajućem poretku za  $s \geq 0$ ,

$$\begin{aligned}\delta_{(1)}(s) &\geq \delta_{(2)}(s) \geq \cdots \geq \delta_{(p)}(s), \\ \gamma_{(1)}^{\rightarrow}(s) &\geq \gamma_{(2)}^{\rightarrow}(s) \geq \cdots \geq \gamma_{(p)}^{\rightarrow}(s), \\ \gamma_{(1)}^{\downarrow}(s) &\geq \gamma_{(2)}^{\downarrow}(s) \geq \cdots \geq \gamma_{(n)}^{\downarrow}(s),\end{aligned}$$

iz skupova

$$\begin{aligned}\{X_{(i),np}^2 v_j(s), i = 1, 2, \dots, p; j = 1, 2, \dots\}, \\ \{D_i^{\rightarrow} v_j(s), i = 1, 2, \dots, p; j = 1, 2, \dots\}, \\ \{D_t^{\downarrow} v_j(s), t = 1, 2, \dots, n; j = 1, 2, \dots\},\end{aligned}$$

respektivno.

### Aproksimacije singularnih vrijednosti

Sljedećim teoremem Davis, Heiny i ostali, u svom radu [9], pružaju vrlo korisne aproksimacije singularnih vrijednosti uzoračke autokovarijacijske matrice linearnog modela definiranog u (3.12).

**Teorem 3.2.2.** *Neka su za linearni model zadan u (3.12) zadovoljeni sljedeći uvjeti:*

- *uvjet regularne varijacije (1.5) za neki  $\alpha \in (0, 4)$ ,*
- *uvjet centriranja  $\mathbb{E}[X] = 0$  ako  $\mathbb{E}[|X|] < \infty$ ,*
- *uvjet sumabilnosti (3.13) za koeficijente matrice  $(h_{kl})$ ,*
- *$(p_n)$  rastu kao u (2.1) za neki  $\beta \geq 0$ .*

Tada za  $s \geq 0$  vrijedi sljedeće:

1. *Uzmimo u obzir dva disjunktna slučaja:  $\alpha \in (0, 2)$  i  $\beta \in (0, \infty)$  ili  $\alpha \in [2, 4)$  i  $\beta$  zadovoljava uvjet 3.7. Tada*

$$a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty. \quad (3.20)$$

2. Neka je  $\beta \in [0, 1]$ . Ako je  $\alpha \in (0, 2]$ ,  $\mathbb{E}[X^2] = \infty$  ili  $\alpha \in [2, 4)$ ,  $\mathbb{E}[X^2] < \infty$  i  $\beta \in (\alpha/2 - 1, 1]$ . Tada

$$a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \gamma_{(i)}^{\rightarrow}(s)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

Neka je  $\beta > 1$ . Ako je  $\alpha \in (0, 2]$ ,  $\mathbb{E}[X^2] = \infty$  ili  $\alpha \in [2, 4)$ ,  $\mathbb{E}[X^2] < \infty$  i  $\beta^{-1} \in (\alpha/2 - 1, 1]$ . Tada

$$a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \gamma_{(i)}^{\downarrow}(s)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

**Napomena 3.2.3.** Dokaz teorema 3.2.2 može se pronaći u radu [14] Heinyja, Mikoscha i Davisa. Drugi dio teorema, uz restriktivnije uvjete na  $(p_n)$ , dokazali su Davis i njegovi kolege u radu [10].

### Konvergencija točkovnih procesa

Uz pomoć teorema 3.2.2 i argumentacije kakvu koriste Davis, Mikosch i Pfaffel u dokazivanju rezultata u radu [10] možemo doći do konvergencije točkovnih procesa normaliziranih singularnih vrijednosti. Prisjetimo se prikaza točaka  $(\Gamma_i)$  jediničnim homogenim Poissonovim procesom iz (3.2). Za  $s \geq 0$  definiramo točkovni proces normaliziranih singularnih vrijednosti s

$$N_n^{\lambda,s} = \sum_{i=1}^p \varepsilon_{a_{np}^{-2}(\lambda_{(i)}(0), \lambda_{(i)}(1), \dots, \lambda_{(i)}(s))}. \quad (3.21)$$

**Teorem 3.2.4.** Pretpostavimo da su zadovoljeni uvjeti teorema 3.2.2. Tada  $(N_n^{\lambda,s})$  konvergira u prostoru točkovnih mjera s prostorom stanja  $(0, \infty)^{s+1}$  kojem je pridružena vague topologija (vidi odjeljak 1.6 za koncept konvergencije slučajnih mjera). Neka je zadovoljeno  $\alpha \in (0, 2]$  i  $\beta \geq 0$  ili  $\alpha \in [2, 4)$ ,  $\mathbb{E}[X^2] < \infty$  i neka je zadovoljen uvjet (3.7). Tada

$$N_n^{\lambda,s} \xrightarrow{d} N = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}(v_j(0), v_j(1), \dots, v_j(s))}, \quad n \rightarrow \infty. \quad (3.22)$$

**Napomena 3.2.5.** Dokaz prethodnog teorema kao i sve pomoće tvrdnje koje se koriste u dokazu se mogu pronaći u radu [9] Davisa, Heinyja i njihovih kolega.



# Poglavlje 4

## Primjene

Rad zaključujemo poglavljem u kojem se nalaze ilustracije teorijskih rezultata iz prethodnih poglavlja na simuliranim i stvarnim podacima te razni primjeri.

### 4.1 Uzoračke kovarijacijske matrice

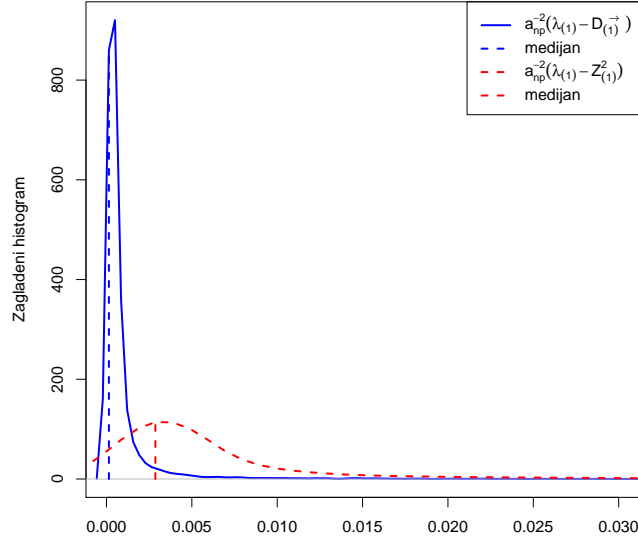
Prisjetimo se, uzoračka kovarijacijska matrica  $\mathbf{Z}_n(0)\mathbf{Z}_n(0)^t = \mathbf{Z}\mathbf{Z}^t$  je pozitivno semidefinitna matrica pa se njezine svojstvene i singularne vrijednosti podudaraju. Nadalje,  $v_j = v_j(0)$ ,  $j \geq 1$ , su svojstvene vrijednosti matrice  $\mathbf{M} = \mathbf{M}(0)$ .

Prva tvrdnja teorema 3.2.2 povlači aproksimaciju uređenih vrijednosti  $(\lambda_{(i)})$  matrice  $\mathbf{Z}\mathbf{Z}^t$  veličinama  $\delta_{(i)}$  koje su dobivene iz uređajnih statistika pridruženih  $(X_{it}^2)$ . Druga tvrdnja teorema nam govori o aproksimaciji vrijednosti  $(\lambda_{(i)})$  veličinama  $(\gamma_{(i)}^{\rightarrow/\downarrow})$  koje su dobivene iz uređajnih statistika pridruženih sumama redaka i stupaca  $(D_i^{\rightarrow/\downarrow})$ . U sljedećem primjeru ilustrirat ćemo kvalitetu ovih aproksimacija.

**Primjer 4.1.1.** *U ovom primjeru komponente matrice  $\mathbf{Z}$  simulirane su iz generalizirane Pareto distribucije, s funkcijom gustoće*

$$f_Z(x) = \begin{cases} \frac{\alpha}{(4|x|)^{\alpha+1}}, & \text{ako } |x| > 1/4 \\ 1, & \text{inače.} \end{cases} \quad (4.1)$$

Po uzoru na Davisa, Heinyja i druge, u [9], simuliramo  $N = 20000$  matrica  $\mathbf{Z}_n$  dimenzije  $p \times n$  za  $p = 200$  i  $n = 1000$  čije komponente imaju funkciju gustoće danu s (4.1). Pretpostavljamo da je  $\beta = 1$ . Uočimo da je  $\mathbf{M} = \mathbf{M}(0)$  ranga jedan i da je  $v_1 = 1$ . Procijenjene distribucije pogrešaka  $a_{np}^{-2}(\lambda_{(1)} - D_{(1)}^{\rightarrow})$  i  $a_{np}^{-2}(\lambda_{(1)} - Z_{(1),np}^2)$  temeljene na simulaciji prikazane su na slici 4.1.



Slika 4.1: Na grafu su prikazane distribucije pogrešaka aproksimacije svojstvenih vrijednosti matrice  $a_{np}^{-2}\mathbf{Z}\mathbf{Z}^t$ . Funkcija gustoće komponenti matrice  $\mathbf{Z}$  dana je s (4.1) za  $\alpha = 1.6$

Na grafu se vidi da je greška aproksimacije vrlo mala, što se podudara s prethodnim teorijskim rezultatima. Vrijedi sljedeća teoretska konvergencija

$$a_{np}^{-2} \sup_i (D_{(i)}^{-} - \lambda_{(i)}) + a_{np}^{-2} \sup_i (Z_{(i),np}^2 - \lambda_{(i)}) \xrightarrow{\mathbb{P}} 0.$$

U slučaju konačnog  $n$ , nizovi  $(D_{(i)}^{-})$  bolje aproksimiraju niz  $(\lambda_{(i)})$ . Nadalje, prema konstrukciji, promatrane razlike bi trebale biti pozitivne, što se i vidi na grafu. Također, na slici 4.1 vidimo da je medijan pogreške aproksimacije  $a_{np}^{-2}(\lambda_{(1)} - D_{(1)}^{-})$  vrlo blizu 0, što ide u prilog točnosti aproksimacije.

Kao izravnu posljedicu teorema 3.2.4 i teorema o neprekidnom preslikavanju dobivamo zajedničku konvergenciju najvećih svojstvenih vrijednosti matrice  $a_{np}^{-2}\mathbf{Z}_n\mathbf{Z}_n^t$  za  $\alpha \in (0, 2)$  i  $\alpha \in (2, 4)$ , kada  $\beta$  zadovoljava uvjet (3.7). Za fiksni  $k \geq 1$  dobivamo

$$a_{np}^{-2}(\lambda_{(1)}, \lambda_{(2)}, \dots, \lambda_{(k)}) \xrightarrow{d} (d_{(1)}, d_{(2)}, \dots, d_{(k)}),$$

gdje su  $d_{(1)} \geq d_{(2)} \geq \dots \geq d_{(k)}$   $k$  najvećih uređenih vrijednosti iz skupa  $\{\Gamma_i^{-2/\alpha} v_j, i = 1, 2, \dots, j = 1, 2, \dots, r\}$ . Za fiksni  $k \geq 1$ , posljedica teorema o neprekidnom preslikavanju

je sljedeća konvergencija o distribuciji:

$$\frac{\lambda_{(1)}}{\lambda_{(1)} + \lambda_{(2)} + \cdots + \lambda_{(k)}} \xrightarrow{d} \frac{d_{(1)}}{d_{(1)} + d_{(2)} + \cdots + d_{(k)}}, \quad n \rightarrow \infty. \quad (4.2)$$

Posljedica primjene teorema o neprekidnom preslikavanju na konvergenciju po distribuciji za točkovne procese iz teorema 3.2.4 je sljedeći korolar. Tvrdnju navodimo bez dokaza, a slični rezultati, kao i sam dokaz, u slučaju  $\alpha \in (2, 4)$  mogu se pronaći u [10].

Prisjetimo se, za  $s \geq 0$   $v_j(s)$  su svojstvene vrijednosti matrice  $\mathbf{M}(s)$ , definirane u (3.18), a  $r(s)$  je rang matrice  $\mathbf{M}(s)$

**Korolar 4.1.2.** *Pretpostavimo da su zadovoljeni uvjeti teorema 3.2.2. Ako je  $\alpha \in (0, 2]$  i  $\mathbb{E}[X^2] = \infty$ , tada*

$$a_{np}^{-2} \left( \lambda_{(1)}, \sum_{i=1}^p \lambda_i \right) \xrightarrow{d} \left( v_1 \Gamma_1^{-2/\alpha}, \sum_{j=1}^r v_j \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha} \right),$$

gdje  $\Gamma_1^{-2/\alpha}$  ima Fréchetovu distribuciju (3.3), a  $\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}$  ima distribuciju pozitivne  $\alpha/2$ -stabilne slučajne varijable. Preciznije, vrijedi:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \xrightarrow{d} \frac{v_1}{\sum_{j=1}^r v_j} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}}, \quad n \rightarrow \infty. \quad (4.3)$$

**Napomena 4.1.3.** *Omjer*

$$\frac{\lambda_{(1)} + \lambda_{(2)} + \cdots + \lambda_{(k)}}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, \quad k \geq 1,$$

igra vrlo važnu ulogu u metodi glavnih komponenti. Taj omjer prikazuje udio varijance u podacima koji možemo objasniti preko prvih  $k$  glavnih komponenti. Iz korolara 4.1.2 slijedi da za fiksni  $k \geq 1$

$$\frac{\lambda_{(1)} + \lambda_{(2)} + \cdots + \lambda_{(k)}}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \xrightarrow{d} \frac{d_{(1)} + d_{(2)} + \cdots + d_{(k)}}{d_{(1)} + d_{(2)} + \cdots}.$$

Općenito, za graničnu varijablu ne znamo točnu formu. Iznimno, u slučaju kada je  $r = 1$  znamo, što je ilustrirano u primjeru 4.1.6. Također, uočimo da se trag matrice  $\mathbf{ZZ}'$  poklapa s  $\lambda_1 + \lambda_2 + \cdots + \lambda_p$ .

U sljedećem primjeru dosadašnje teorijske rezultate ilustrirat ćemo na takozvanom modelu pomičnih prosjeka (eng. *moving average model*).

**Primjer 4.1.4.** *Pretpostavimo da je  $\alpha \in (0, 2)$  i*

$$Z_{it} = X_{it} + X_{i,t-1} - 2(X_{i-1,t} - X_{i-1,t-1}), \quad i, t \in \mathbb{Z}. \quad (4.4)$$

*Uočimo, u ovom slučaju elementi matrice  $\mathbf{H}$  koji su različiti od 0 su*

$$h_{00} = 1, h_{01} = 1, h_{10} = -2 \quad i \quad h_{11} = 2.$$

*Iz toga slijedi da  $\mathbf{M} = \mathbf{H}\mathbf{H}^t$  ima pozitivne svojstvene vrijednosti  $v_1 = 8$  i  $v_2 = 2$ . Uz prethodno izračunate vrijednosti, granični proces iz (3.22) je u ovom primjer dan relacijom*

$$N = \sum_{i=1}^{\infty} \varepsilon_{8\Gamma_i^{-2/\alpha}} + \sum_{i=1}^{\infty} \varepsilon_{2\Gamma_i^{-2/\alpha}},$$

*tako da*

$$a_{np}^{-2}(\lambda_{(1)}, \lambda_{(2)}) \xrightarrow{d} (8\Gamma_1^{-2/\alpha}, \min(2\Gamma_1^{-2/\alpha}, 8\Gamma_2^{-2/\alpha})).$$

*Činjenicu da  $U = \Gamma_1/\Gamma_2$  ima uniformnu distribuciju na  $(0, 1)$  možemo iskoristiti za sljedeći račun*

$$\mathbb{P}(2\Gamma_1^{-2/\alpha} > 8\Gamma_2^{-2/\alpha}) = \mathbb{P}(\Gamma_1/\Gamma_2 < 2^{-\alpha}) = 2^{-\alpha} \in (1/4, 1).$$

*Preciznije, za normaliziranu spektralnu razliku vrijedi*

$$a_{np}^{-2}(\lambda_{(1)} - \lambda_{(2)}) \xrightarrow{d} 6\Gamma_1^{-2/\alpha} \mathbf{1}_{\{\Gamma_1 4^{\alpha/2} < \Gamma_2\}} + 8(\Gamma_1^{-2/\alpha} - \Gamma_2^{-2/\alpha}) \mathbf{1}_{\{\Gamma_1 4^{\alpha/2} > \Gamma_2\}}$$

*a za samonormalizirajuću spektralnu razliku*

$$\begin{aligned} \frac{\lambda_{(1)} - \lambda_{(2)}}{\lambda_{(1)}} &\xrightarrow{d} \frac{6}{8} \mathbf{1}_{\{\Gamma_1 2^{\alpha} < \Gamma_2\}} + (1 - (\Gamma_1/\Gamma_2)^{2/\alpha}) \mathbf{1}_{\{\Gamma_1 2^{\alpha} > \Gamma_2\}} \\ &= \frac{3}{4} \mathbf{1}_{\{U 2^{\alpha} < 1\}} + (1 - U^{2/\alpha}) \mathbf{1}_{\{U 2^{\alpha} > 1\}} = Y. \end{aligned}$$

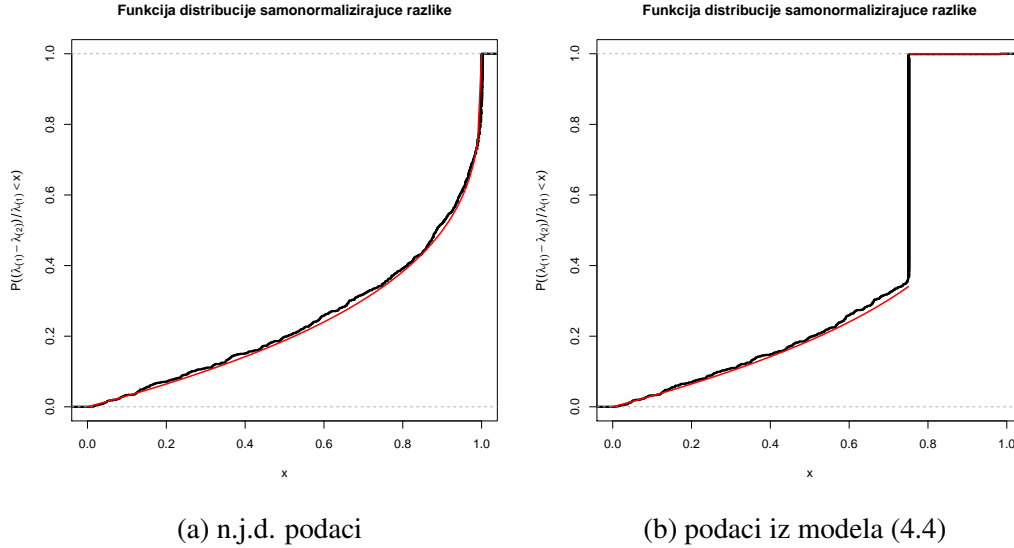
*Granična distribucija spektralne razlike ima atom u  $3/4$  s vjerojatnošću  $2^{-\alpha}$ , tj.  $\mathbb{P}(Y = 3/4) = 2^{-\alpha}$  i*

$$\mathbb{P}(Y \leq x) = 1 - (1 - x)^{\alpha/2}, \quad x \in (0, 3/4).$$

*U slučaju kada su komponente granične distribucije nezavisne i jednako distribuirane, samonormalizirajuća spektralna razlika ima funkciju distribucije*

$$F(x) = 1 - (1 - x)^{\alpha/2}, \quad x \in [0, 1].$$

*Dakle, atom nestaje u slučaju nezavisno jednako distribuiranih komponenti. Na slici 4.2 uspoređene su funkcija distribucije slučajne varijable  $Y$  s  $F$  za  $\alpha = 0.6$ . Jasno se vidi atom u točki  $3/4$ .*



Slika 4.2: Funkcije distribucije  $(\lambda_{(1)} - \lambda_{(2)})/\lambda_{(1)}$  za n.j.d. podatke (lijeva slika) i podatke generirane iz modela (4.4) (desna slika). Na svakom grafu uspoređene su procijenjena uzoračka funkcija distribucije (crna linija) i teoretska krivulja (crvena linija), na temelju 1000 simulacija matrica dimenzija  $200 \times 1000$  s distribucijom komponentata matrice kao u (4.1).

Između ostalog, vrijedi i sljedeća konvergencija po distribuciji

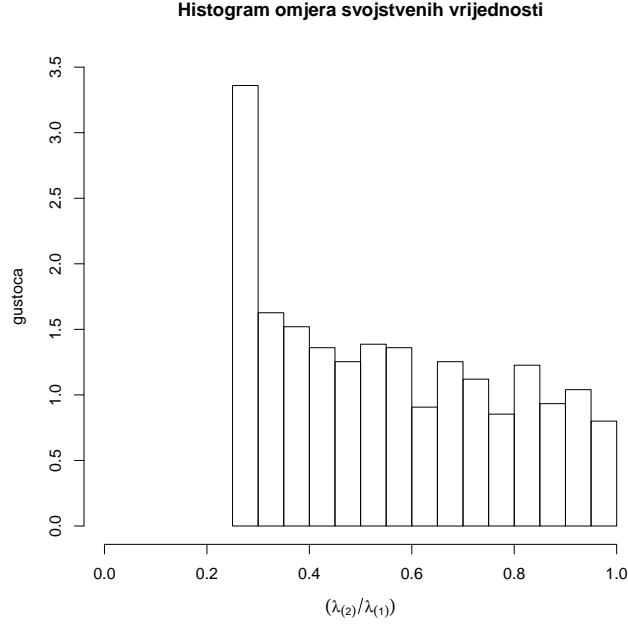
$$(a_{np}^{-2}\lambda_{(1)}, \lambda_{(2)}/\lambda_{(1)}) \xrightarrow{d} (8\Gamma_1^{-2/\alpha}, \frac{1}{4}\mathbf{1}_{\{U < 2^{-\alpha}\}} + (1 - U^{2/\alpha})\mathbf{1}_{\{U \geq 2^{-\alpha}\}}).$$

Iz prethodne relacije vidimo da je nosač granične distribucije druge varijable  $[1/4, 1)$ , te da ima masu  $2^{-\alpha}$  u točki  $1/4$ . Na slici 4.3 prikazan je histogram omjera  $(\lambda_{(2)}/\lambda_{(1)})^{2/\alpha}$  koji je dobiven na temelju 1000 simulacija iz modela (4.4) dimenzija  $p = 200$  i  $n = 1000$ , pri čemu je distribucija šuma studentova  $t$ -distribucija s  $\alpha = 1.5$  stupnjeva slobode. Uočimo  $2^{-\alpha} = 0.35355\dots$  te da je histogram izuzetno blizak onome što bismo očekivali dobiti na temelju uzorka iz distribucije  $2^{-\alpha}\mathbf{1}_{\{U < 2^{-\alpha}\}} + U\mathbf{1}_{\{U \geq 2^{-\alpha}\}}$ .

Masa granične diskretne komponente omjera može biti mnogo veća uz drugačije uvjete na veličinu  $a_{np}^{-2}\lambda_{(1)}$ . Preciznije, za svaki  $\varepsilon \in (0, 1/4)$  i  $x > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\varepsilon < \lambda_{(2)}/\lambda_{(1)} \leq 1/4 | \lambda_{(1)} > a_{np}^2 x) = \mathbb{P}(\Gamma_1/\Gamma_2 \leq (x/8)^{-\alpha/2}) = G(x).$$

Funkcija  $G$  se približava 1 kada  $x \rightarrow \infty$  što možemo shvatiti kao indikator brzine kojom dvije najveće svojstvene vrijednosti postaju linearno zavisne. Nadalje, iz napomene 4.1.3



Slika 4.3: Histogram dobiven na temelju simulacije 1000 omjera  $(\lambda_{(2)}/\lambda_{(1)})^{2/\alpha}$  iz modela (4.4).

slijedi

$$\frac{\lambda_{(1)}}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \xrightarrow{d} \frac{4}{5} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}}.$$

Iz prethodne relacije vidimo da je granična slučajna varijabla stohastički manja nego u slučaju nezavisno jednako distribuiranih komponenti (vidi 4.3).

**Primjer 4.1.5.** U prethodnim primjerima ilustrirano je ponašanje dvije najveće svojstvene vrijednosti u općenitom slučaju kada je rang  $r$  matrice  $\mathbf{M}$  veći od jedan. Općenito, vrijedi:

$$\frac{\lambda_{(2)}}{\lambda_{(1)}} \xrightarrow{d} \frac{v_2}{v_1} \mathbf{1}_{\{U < (v_2/v_1)^{\alpha/2}\}} + U^{2/\alpha} \mathbf{1}_{\{U \geq (v_2/v_1)^{\alpha/2}\}}.$$

Preciznije, granična samonormalizirajuća spektralna razlika može se prikazati u sljedećem obliku:

$$\frac{\lambda_{(1)} - \lambda_{(2)}}{\lambda_{(1)}} \xrightarrow{d} \frac{v_1 - v_2}{v_1} \mathbf{1}_{\{U < (v_2/v_1)^{\alpha/2}\}} + (1 - U^{2/\alpha}) \mathbf{1}_{\{U \geq (v_2/v_1)^{\alpha/2}\}}.$$

Granična varijabla poprima vrijednosti u  $(0, 1 - v_2/v_1]$  te ima atom u lijevoj rubnoj točki. Ovo je u suprotnosti od slučaja n.j.d. komponenti kada je  $r = 1$  (stoga  $v_2 = 0$ ) uključujući

slučaj kada su redovi nezavisno jednako distribuirani te rastavljiv model (vidi primjer 4.1.6).

**Primjer 4.1.6.** Smatramo da je model rastavljiv ako je  $h_{kl} = \theta_k c_l$ ,  $k, l \in \mathbb{Z}$ , gdje su  $(\theta_k)$  i  $(c_l)$  nizovi realnih brojeva takvi da su zadovoljeni uvjeti teorema 3.2.2. U ovom slučaju

$$\mathbf{M} = \sum_{l \in \mathbb{Z}} c_l^2 (\theta_i \theta_j)_{i,j \in \mathbb{Z}}.$$

Uočimo,  $r = 1$  sa samo jednom nenegativnom svojstvenom vrijednosti

$$v_1 = \sum_{l \in \mathbb{Z}} c_l^2 \sum_{k \in \mathbb{Z}} \theta_k^2.$$

U ovom slučaju, granični točkovni proces iz teorema 3.2.4 je Poissonova slučajna mjera na  $(0, \infty)$  za koju je mjera očekivanja od  $(y, \infty)$  dana s  $(v_1/y)^{\alpha/2}$ ,  $y > 0$  (vidi [9] za više). Normalizirane svojstvene vrijednosti se slično ponašaju kao u slučaju nezavisno jednako distribuiranih komponenti matrice podataka. Na primjer, log-razlike poprimaju iste granične vrijednosti kao u n.j.d. slučaju za fiksni  $k \geq 1$ ,

$$(\log \lambda_{(1)} - \log \lambda_{(2)}, \dots, \log \lambda_{(k)} - \log \lambda_{(k+1)}) \xrightarrow{d} -\frac{2}{\alpha} (\log(\Gamma_1/\Gamma_2), \dots, \Gamma_k/\Gamma_{k+1}).$$

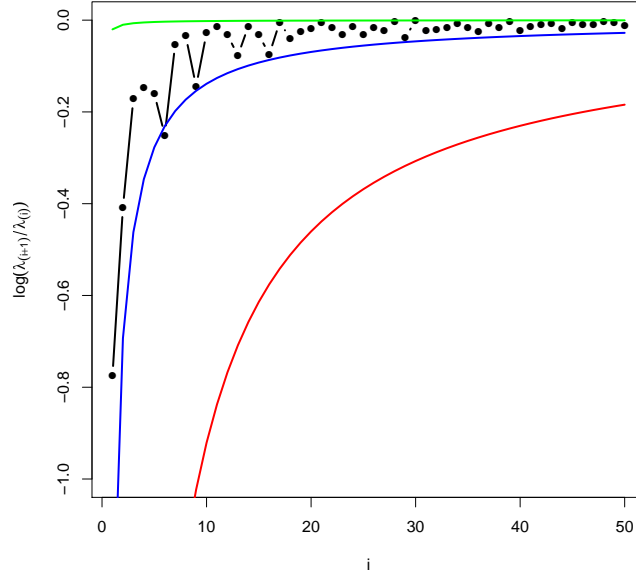
Isto opažanje možemo primjeniti i na omjer najveće svojstvene vrijednosti i traga matrice  $\mathbf{Z}\mathbf{Z}'$ , u slučaju kada je  $\alpha \in (0, 2)$  :

$$\frac{\lambda_{(1)}}{\text{tr}(\mathbf{Z}\mathbf{Z}')} = \frac{\lambda_{(1)}}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \xrightarrow{d} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}}.$$

Granična distribucija samonormalizirajuće spektralne razlike nema atom kao ni u slučaju n.j.d. komponenti matrice podataka.

## 4.2 Analiza podataka indeksa S&P 500

U ovom odjeljku napraviti ćemo kratku analizu najveće svojstvene vrijednosti univarijatnih vremenskih nizova log-povrata koji tvore dionički indeks S&P 500 (na stranici 28 nalazi se opis podataka). Unatoč empirijskim dokazima da ovi univarijatni vremenski nizovi imaju teške repove (vidi sliku 3.3) ne očekujemo da imaju iste repne indekse. Jedan način da rješimo ovaj problem je ignoriranje činjenice da su repni indeksi različiti uz opravdanje da su razlike među indeksima male po iznosu te da dolaze iz greške u obradi velike količine podataka i same procjene veličina. Drugi način bi bio da agregiramo vremenske nizove sa



Slika 4.4: Logaritmi omjera svojstvenih vrijednosti  $\lambda_{(i+1)}/\lambda_{(i)}$  za vremenske nizove koji tvore S&P 500 indeks, nakon transformacije rangova. Na grafu su također prikazani 1, 50, 90%-tni kvantili (gornja-zelena, srednja-plava i donja-crvena linija, respektivno) varijabli  $\log((\Gamma_i/\Gamma_{i+1})^2)$ .

sličnim repnim indeksom u iste grupe. Uočimo, u tom slučaju dimenzija  $p$  bi se smanjila. Nadalje, grupiranje bi proveli proizvoljnom klasifikacijskom metodom. U ovoj analizi, po uzoru na Davisa, Heinyja i ostale u [9], problem ćemo riješiti preko transformacije rangova. Prednost ovog pristupa rješavanju problema je ciljano standardiziranje repova, dok je mana gubitak kovarijacijske strukture nastale transformacijom podataka. Za danu matricu podataka  $(R_{it})_{i=1,\dots,p;t=1,\dots,n}$  konstruiramo matricu  $\mathbf{Z}$  transformacijom rangova

$$Z_{it} = -\left[ \log \left( \frac{1}{n+1} \sum_{\tau=1}^n \mathbf{1}_{\{R_{it} \leq R_{i\tau}\}} \right) \right]^{-1}, \quad i = 1, \dots, p; t = 1, \dots, n.$$

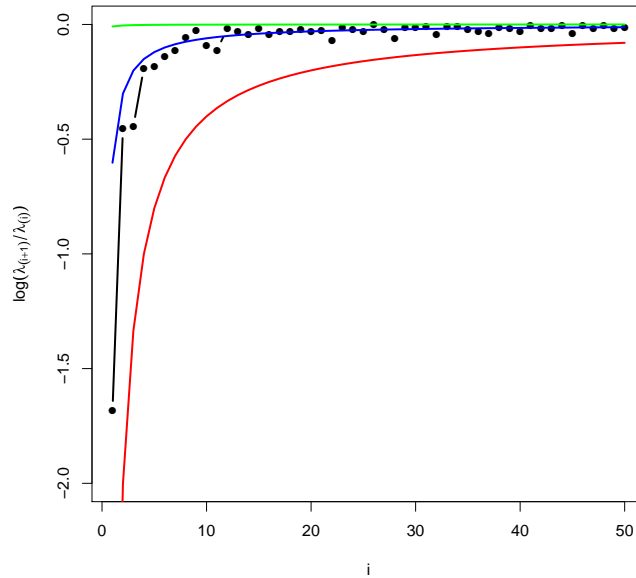
Kada bi redovi  $R_{i1}, \dots, R_{in}$  bili nezavisno jednako distribuirani (ili, općenitije, ergodični i stacionarni) s neprekidnom distribucijom, tada bi prosjek unutar logaritamske funkcije bio asimptotski uniforman na  $(0, 1)$  kako  $n \rightarrow \infty$ . Stoga bi  $Z_{it}$  imao asimptotski standardnu Fréchetovu  $\Phi_1$  distribuciju. U nastavku pretpostavljamo da su ranije spomenuti univarijatni vremenski nizovi log-povrata S&P 500 indeksa podvrgnuti odgovarajućoj transformaciji rangova te marginalnu distribuciju blizu Fréchetove  $\Phi_1$ . U nastavku ćemo također koristiti oznaku  $\mathbf{Z}$  za transformiranu matricu.



Na slici 4.4 su prikazani uzastopni logaritmi omjera padajuće uređenih svojstvenih vrijednosti  $\log(\lambda_{(i+1)}/\lambda_{(i)})$  matrice  $\mathbf{ZZ}^T$ . Na grafu možemo uočiti da su omjeri mali po iznosu čak i za male  $i$ , tj. najveće svojstvene vrijednosti. Nadalje, na slici su također prikazani kvantili slučajnih varijabli  $\log((\Gamma_i/\Gamma_{i+1})^2)$  koji su dobiveni iz formule

$$\mathbb{P}((\Gamma_i/\Gamma_{i+1})^{2/\alpha} \leq x) = x^{i\alpha/2}, \quad x \in (0, 1). \quad (4.5)$$

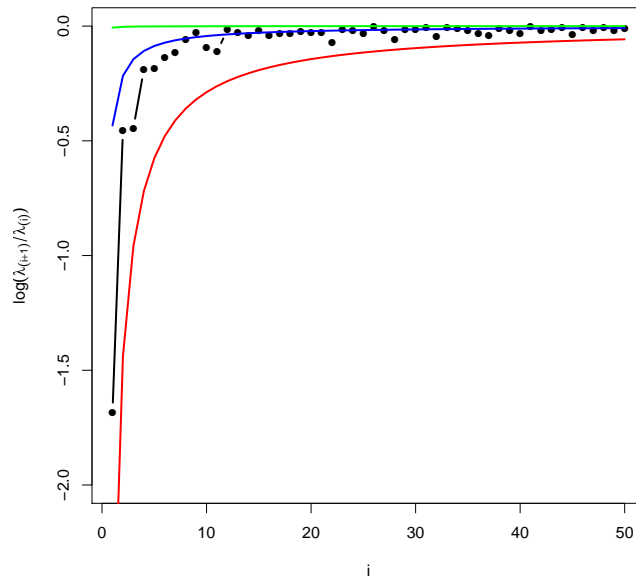
Kako  $i$  raste, distribucija je koncentrirana oko 1 (na grafu su prikazane logaritamske vrijednosti, stoga je na grafu koncentrirana oko 0), što je u skladu s *Jakim zakonom velikih brojeva* koji povlači  $\Gamma_i/\Gamma_{i+1} \xrightarrow{g.s.} 1$ , kako  $i \rightarrow \infty$ . Asimptotske distribucije u (4.5) odgovaraju slučaju kada je rang matrice  $\mathbf{M}$  jednak jedan. Tada su uključeni i slučaj kada su komponente nezavisne i jednako distribuirane i rastavljiv model (vidi primjer 4.1.6).



Slika 4.5: Prikaz logaritama omjera svojstvenih vrijednosti  $\lambda_{(i+1)}/\lambda_{(i)}$  za vremenske nizove koji tvore S&P 500 indeks za originalne podatke. Na grafu su također prikazani 1, 50, 90%-tni kvantili (gornja-zelena, srednja-plava i donja-crvena linija, redom) varijabli  $\log((\Gamma_i/\Gamma_{i+1})^{2/2.3})$ . Vidi sliku 4.4 za usporedbu.

Za usporedbu, na slici 4.5 su prikazani logaritmi omjera svojstvenih vrijednosti za ne-transformiranu matricu podataka S&P 500 indeksa. Kao i na prethodnoj slici, uz omjere svojstvenih vrijednosti prikazani su i 1, 50, 99%-tni kvantili varijabli  $\log((\Gamma_i/\Gamma_{i+1})^{2/\alpha})$ . Po uzoru na Davisa, Heinyja i ostale u [9], te motivacije iz slike 3.3 na kojoj se nalaze procijenjeni repni indeksi za podatke, prikazani kvantili su izračunati iz (4.5) za  $\alpha = 2.3$ .

Odabirom  $\alpha = 2.3$  pretpostavljamo jednakost repnih indeksa log-povrata, a odabiremo na način da uzmemo približno najmanji repni indeks, tj. uzimamo približno onaj za koji je 99% ostalih repnih indeksa veće. Grafovi na slici 4.4 i na slici 4.5 su vrlo slični. Ako uzmemo u obzir različite vrijednosti na y osi, možemo uočiti da graf na slici 4.5 za netransformirane podatke ima veće omjere svojstvenih vrijednosti za veće  $i$ , tj. omjeri manjih svojstvenih vrijednosti su malo veći od onih za podatke nakon transformacije rangova. Dodatno, radi provjere robusnosti rezultata konstruiramo kvantile za drugačije odabrani  $\alpha$ . Za novi  $\alpha$  uzimamo aritmetičku sredinu svih repnih indeksa prikazanih na slici 3.3, pri čemu za svaki vremenski niz log-povrata uzimamo aritmetičku sredinu lijevog i desnog repnog indeksa. Na slici 4.6 prikazani su omjeri svojstvenih vrijednosti za originalne podatke uz  $\alpha = 3.2$ . Uspoređujući slike 4.5 i 4.6 ne uočavamo značajnu razliku, pa zaključujemo da je konstrukcija robusna na odabir  $\alpha$ .



Slika 4.6: Prikaz logaritama omjera svojstvenih vrijednosti  $\lambda_{(i+1)}/\lambda_{(i)}$  za vremenske nizove koji tvore S&P 500 indeks za originalne podatke. Na grafu su također prikazani 1, 50, 90%–tni kvantili (gornja-zelena, srednja-plava i donja-crvena linija, redom) varijabli  $\log((\Gamma_i/\Gamma_{i+1})^{2/3.2})$ . Vidi sliku 4.5 za usporedbu.

### 4.3 Sume kvadrata uzoračke autokovarijacijske matrice

U ovom odjeljku promatramo funkcije kvadratnih vrijednosti funkcija  $\mathbf{A}_n(s) = \mathbf{Z}_n(0)\mathbf{Z}_n(s)^t$  dane s  $\mathbf{A}_n(s)\mathbf{A}_n(s)^t$  za  $s = 0, 1, \dots$ . Po definiciji singularnih vrijednosti matrice (vidi (3.16)), svojstvene vrijednosti pozitivno semidefinitne matrice  $\mathbf{A}_n(s)\mathbf{A}_n(s)^t$  su  $(\lambda_i^2(s))_{i=1,2,\dots,p}$ .

Sljedeći rezultat je posljedica teorema 3.2.2.

**Propozicija 4.3.1.** *Neka su za linearni proces (3.12) zadovoljeni uvjeti teorema 3.2.2. Tada, za  $s \geq 0$  vrijede sljedeće tvrdnje:*

1. *Uzmimo u obzir dva disjunktna slučaja:  $\alpha \in (0, 2)$  i  $\beta \in (0, \infty)$  ili  $\alpha \in [2, 4)$  i  $\beta$  zadovoljava uvjet 3.7. Tada*

$$a_{np}^{-4} \max_{i=1,2,\dots,p} |\lambda_{(i)}^2(s) - \delta_{(i)}^2(s)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

2. *Neka je  $\beta \in [0, 1]$ . Ako je  $\alpha \in (0, 2]$ ,  $\mathbb{E}[X^2] = \infty$  ili  $\alpha \in [2, 4)$ ,  $\mathbb{E}[X^2] < \infty$  i  $\beta \in (\alpha/2 - 1, 1]$ . Tada*

$$a_{np}^{-4} \max_{i=1,2,\dots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\rightarrow}(s))^2| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

*Neka je  $\beta > 1$ . Ako je  $\alpha \in (0, 2]$ ,  $\mathbb{E}[X^2] = \infty$  ili  $\alpha \in [2, 4)$ ,  $\mathbb{E}[X^2] < \infty$  i  $\beta^{-1} \in (\alpha/2 - 1, 1]$ . Tada*

$$a_{np}^{-4} \max_{i=1,2,\dots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\downarrow}(s))^2| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

*Dokaz.* Dokaz tvrdnje (1). Ako pokažemo da vrijedi

$$a_{np}^{-2} \max_{i=1,2,\dots,p} (\lambda_{(i)}(s) + \delta_{(i)}(s)) = O_{\mathbb{P}}(1)$$

tvrdnja će slijediti iz teorema 3.2.2. Po teoremu 3.2.4 vrijedi

$$a_{np}^{-2} \max_{i=1,2,\dots,p} \lambda_{(i)}(s) = a_{np}^{-2} \lambda_{(1)}(s) \xrightarrow{d} c\xi_{\alpha/2}, \quad (4.6)$$

gdje  $\xi_{\alpha/2}$  ima Fréchetovu  $\Phi_{\alpha/2}$  distribuciju. Također, po teoremu 3.2.2 vrijedi

$$a_{np}^{-2} \max_{i=1,2,\dots,p} \delta_{(i)}(s) \xrightarrow{d} c\xi_{\alpha/2}.$$

Sada, koristeći opet prvu tvrdnju teorema 3.2.2 vrijedi

$$\begin{aligned} & a_{np}^{-4} \max_{i=1,2,\dots,p} |\lambda_{(i)}^2(s) - \delta_{(i)}^2(s)| \\ & \leq [a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)|] [a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) + \delta_{(i)}(s)|] \\ & \leq [a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)|] [a_{np}^{-2} \max_{i=1,2,\dots,p} (|\lambda_{(i)}(s)| + |\delta_{(i)}(s)|)] \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty. \end{aligned}$$

Time je prva tvrdnja teorema dokazana.

Dokaz tvrdnje (2). Pretpostavimo da je sada  $\beta \in [0, 1]$  i  $\alpha \in (0, 2]$ ,  $\mathbb{E}[X^2] = \infty$  ili  $\alpha \in [2, 4)$ ,  $\mathbb{E}[X^2] < \infty$  i  $\beta \in (\alpha/2 - 1, 1]$ . Tada (4.6) također vrijedi i prema drugoj tvrdnji teorema 3.2.2 i teoremu 3.2.4 vrijedi

$$a_{np}^{-2} \max_{i=1,2,\dots,p} \gamma_{(i)}^{\rightarrow}(s) \xrightarrow{d} c\xi_{\alpha/2}.$$

Iz prethodno navedenog, sada vrijedi

$$\begin{aligned} & a_{np}^{-4} \max_{i=1,2,\dots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\rightarrow}(s))^2| \\ & \leq [a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \gamma_{(i)}^{\rightarrow}(s)|] [a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) + \gamma_{(i)}^{\rightarrow}(s)|] \\ & \leq [a_{np}^{-2} \max_{i=1,2,\dots,p} |\lambda_{(i)}(s) - \gamma_{(i)}^{\rightarrow}(s)|] [a_{np}^{-2} \max_{i=1,2,\dots,p} (|\lambda_{(i)}(s)| + |\gamma_{(i)}^{\rightarrow}(s)|)] \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty. \end{aligned}$$

Dokaz zadnje tvrdnje je analogan kao dokaz prethodne.  $\square$

Korištenjem propozicije 4.3.1 i teorema o neprekidnom preslikavanju može se pokazati sljedeći rezultat za svojstvene vrijednosti

$$w_{(1)}(s_0, s_1) \geq \dots \geq w_{(p)}(s_0, s_1), \quad 0 \leq s_0 \leq s_1,$$

pozitivno semidefinitne matrice

$$\sum_{s_0=1}^{s_1} \mathbf{A}_n(s) \mathbf{A}_n(s)^t. \quad (4.7)$$

**Propozicija 4.3.2.** Pretpostavimo da je  $0 \leq s_0 \leq s_1$  i da su zadovoljeni uvjeti teorema 3.2.2. Ako su  $\alpha \in (0, 4)$  i  $\beta \in (0, 2] \cap (\alpha/2 - 1, 1]$  tada vrijedi

$$a_{np}^{-4} |w_{(i)}(s_0, s_1) - \omega_{(i)}(s_0, s_1)| \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty,$$

gdje su  $\omega_{(i)}(s_0, s_1)$  padajuće uređene vrijednosti iz skupa  $\{X_{(i),np}^4 v_j(s_0, s_1), i = 1, \dots, p; j = 1, 2, \dots\}$  i  $(v_j(s_0, s_1))$  su padajuće uređene svojstvene vrijednosti matrice  $\sum_{s=s_0}^{s_1} \mathbf{M}(s) \mathbf{M}(s)^t$ .

**Primjer 4.3.3.** *Prisjetimo se rastavljivog modela iz primjera 4.1.6, tj.  $h_{kl} = \theta_k c_l$ ,  $k, l \geq 0$ , gdje su  $(\theta_k)$  i  $(c_l)$  nizovi realnih brojeva takvi da su uvjeti teorema 3.2.2 za  $(h_{kl})$  zadovoljeni. U nastavku koristimo oznaku  $\Theta_{ij} = \theta_i \theta_j$ .  $\Theta$  je simetrična matrica ranga 1 i jedina pripadna svojstvena vrijednost različita od nule jednaka je  $\gamma_\theta(s) = \sum_{k=0}^{\infty} \theta_k^2$ . Stoga, možemo zaključiti da je pozitivno semidefinitna. Iz (3.17) slijedi*

$$\mathbf{M}(s) = \gamma_c(s)\Theta, \quad s \geq 0.$$

gdje

$$\gamma_c(s) = \sum_{l=0}^{\infty} c_l c_{l+s}, \quad s \geq 0.$$

Matrica  $\mathbf{M}(s)$  ima samo jednu svojstvenu vrijednost različitu od nule i ona je jednaka  $\gamma_c(s)\gamma_\theta(0)$ . Brojevi  $(\gamma_c(s))$  mogu biti pozitivni i negativni (oni tvore autokovarijacijsku funkciju stacionarnog linearnog procesa s koeficijentima  $(c_l)$ ). Prema tome, matrica  $\mathbf{M}(s)$  je pozitivno semidefinitna ili negativno semidefinitna. Pretpostavimo da je sljedeća matrica pozitivno semidefinitna

$$\sum_{s=s_0}^{s_1} \mathbf{M}(s)\mathbf{M}(s)^t = \sum_{s=s_0}^{s_1} \gamma_c^2(s)\Theta\Theta^t.$$

Ova matrica je ranga 1 i najveća pripadna svojstvena vrijednost dana je sa

$$C_{c,\theta}(s_0, s_1) = \sum_{s=s_0}^{s_1} \gamma_c^2(s)\gamma_\theta^2(s).$$

Propozicija 4.3.2 povlači da su padajuće uređene svojstvene vrijednosti matrice  $a_{np}^{-4} \sum_{s_0=1}^{s_1} \mathbf{A}_n(s)\mathbf{A}_n(s)^t$  uniformno aproksimirane veličinama

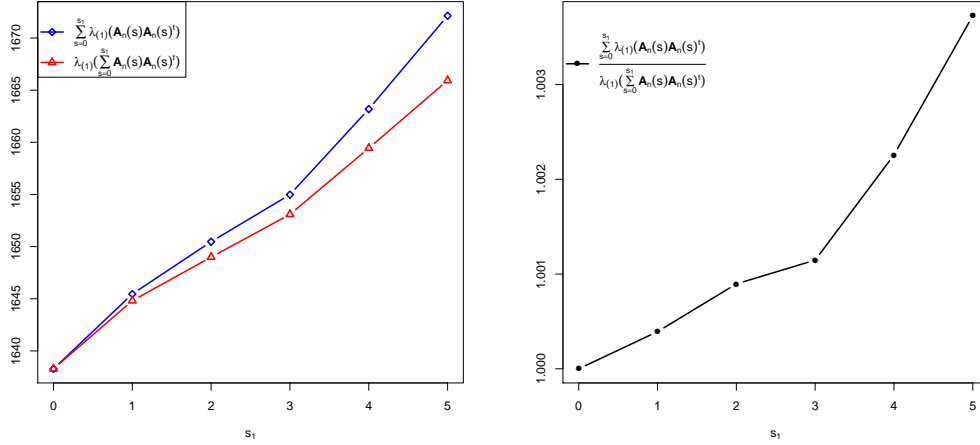
$$a_{np}^{-4} X_{(i),np}^4 C_{c,\theta}(s_0, s_1), \quad i = 1, \dots, p. \quad (4.8)$$

Zbog

$$C_{c,\theta}(s_0, s_1) = \sum_{i=s_0}^{s_1} C_{c,\theta}(i, i)$$

dolazimo do iznimnog rezultata

$$\begin{aligned} & a_{np}^{-4} \max_{i=1,2,\dots,p} \left| \lambda_{(i)} \left( \sum_{s_0=1}^{s_1} \mathbf{A}_n(s)\mathbf{A}_n(s)^t \right) - X_{(i),np}^4 C_{c,\theta}(s_0, s_1) \right| \\ &= a_{np}^{-4} \max_{i=1,2,\dots,p} \left| \sum_{s_0=1}^{s_1} \lambda_{(i)} (\mathbf{A}_n(s)\mathbf{A}_n(s)^t) - X_{(i),np}^4 C_{c,\theta}(s_0, s_1) \right| + \mathbf{O}_{\mathbb{P}}(1). \end{aligned}$$



Slika 4.7: Najveće svojstvene vrijednosti sume kvadrata autokovarijacijskih matrica uspoređene sa sumom najvećih svojstvenih vrijednosti istih matrica za matricu podataka S&P 500 dioničkog indeksa za različite vrijednosti  $s_1$ . Vrijednosti su vrlo bliske. Također, prikazani su omjeri dviju vrijednosti.

Za  $s_1 \geq s_0$  dobivamo konvergenciju po distribuciji točkovnih procesa prema Poissonovoj slučajnoj mjeri (vidi odjeljak 1.6 za koncept konvergencije):

$$\sum_{i=1}^p \mathcal{E} \left( a_{np}^{-4} \left( \lambda_i \left( \sum_{s_0=1}^{s_0} \mathbf{A}_n(s) \mathbf{A}_n(s)^t \right), \dots, \lambda_i \left( \sum_{s_0=1}^{s_1} \mathbf{A}_n(s) \mathbf{A}_n(s)^t \right) \right) \right) \\ \rightarrow \sum_{i=1}^{\infty} \mathcal{E}_{\Gamma_i^{-4/\alpha}} (C_{c,\theta}(s_0, s_0), \dots, C_{c,\theta}(s_0, s_1)), \quad n \rightarrow \infty.$$

**Primjer 4.3.4.** Na slici 4.7 uspoređene su najveće svojstvene vrijednosti matrice  $\lambda_{(1)}(\sum_{s=0}^{s_1} \mathbf{A}_n(s) \mathbf{A}_n(s)^t)$  za  $s_1 = 0, 1, 2, \dots, 5$  i  $\sum_{s=0}^{s_1} \lambda_{(1)}(\mathbf{A}_n(s) \mathbf{A}_n(s)^t)$  za vremenske nizove log-povrata S&P 500 dioničkog indeksa (podaci su opisani na stranici 28). Podaci nisu podvrgnuti transformaciji rangova. Na lijevom grafu možemo uočiti da su veličine vrlo blizu za različite vrijednosti  $s_1$ . Ovu pojavu možemo objasniti strukturom svojstvenih vrijednosti opisanoj u primjeru 4.3.3. Također, uočimo da najveća svojstvena vrijednost matrice  $\mathbf{A}_n(0) \mathbf{A}_n(0)^t$  najviše doprinosi veličinom vrijednostima na slici 4.7.

# Bibliografija

- [1] Naum Ilić Akhiezer. *The classical moment problem: and some related questions in analysis*, volume 5. Oliver & Boyd, 1965.
- [2] Zhi-Dong Bai. Methodologies in spectral analysis of large dimensional matrices, a review. *Statistica Sinica*, 9:611–677, 1999.
- [3] Zhi-Dong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.
- [4] Zhi-Dong Bai and Yong-Qua Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Annals of probability*, 21(3):1275–1294, 1993.
- [5] Serban Belinschi, Amir Dembo, and Alice Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Communications in Mathematical Physics*, 289(3):1023–1055, 2009.
- [6] Florent Benaych-Georges and Sandrine Péché. Localization and delocalization for heavy tailed band matrices. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 50, pages 1385–1403. Institut Henri Poincaré, 2014.
- [7] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 1997.
- [8] Nicholas H. Bingham, Charles M. Goldie, and Jef L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [9] Richard A. Davis, Johannes Heiny, Thomas Mikosch, and Xiaolei Xie. Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes*, 19(3):517–547, 2016.
- [10] Richard A. Davis, Thomas Mikosch, and Oliver Pfaffel. Asymptotic theory for the sample covariance matrix of a heavy-tailed multivariate time series. *Stochastic Processes and their Applications*, 126(3):767–799, 2016.

- [11] Richard A. Davis, Oliver Pfaffel, and Robert Stelzer. Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. *Stochastic Processes and their Applications*, 124(1):18–50, 2014.
- [12] Stuart Geman. A limit theorem for the norm of random matrices. *The Annals of Probability*, pages 252–261, 1980.
- [13] Johannes Heiny and Thomas Mikosch. Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. *Stochastic Processes and their Applications*, 127(7):2179–2207, 2017.
- [14] Johannes Heiny, Thomas Mikosch, and Richard A. Davis. Limit theory for the singular values of the sample autocovariance matrix function of multivariate time series. *Work in progress*, 2015.
- [15] Tiefeng Jiang. The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā: The Indian Journal of Statistics*, pages 35–48, 2004.
- [16] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.
- [17] Thomas Mikosch. *Non-life insurance mathematics: an introduction with the Poisson process*. Springer Science & Business Media, 2009.
- [18] Sean O’Rourke, Van Vu, and Ke Wang. Eigenvectors of random matrices: a survey. *Journal of Combinatorial Theory, Series A*, 144:361–442, 2016.
- [19] Debashis Paul and Alexander Aue. Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150:1–29, 2014.
- [20] Sidney I. Resnick. *Extreme values, regular variation and point processes*. Springer.
- [21] Sidney I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media, 2007.
- [22] Alexander Soshnikov. Poisson statistics for the largest eigenvalues of wigner random matrices with heavy tails. *Electronic Communications in Probability*, 9:82–91, 2004.
- [23] Alexander Soshnikov. Poisson statistics for the largest eigenvalues in random matrix ensembles. In *Mathematical physics of quantum mechanics*, pages 351–364. Springer, 2006.
- [24] Terence Tao and Van Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, 298(2):549–572, 2010.



- [25] Terence Tao and Van Vu. Random covariance matrices: Universality of local statistics of eigenvalues. *The Annals of Probability*, pages 1285–1315, 2012.
- [26] Konstantin Tikhomirov. The limit of the smallest singular value of random matrices with iid entries. *Advances in Mathematics*, 284:1–20, 2015.
- [27] Craig A. Tracy and Harold Widom. Distribution functions for largest eigenvalues and their applications. In *Proceedings of the International Congress of Mathematicians, (Beijing, 2002)*, 1:587–596, 2002.
- [28] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20(1/2):32–52, 1928.
- [29] Han Xiao and Wang Zhou. Almost sure limit of the smallest eigenvalue of some sample correlation matrices. *Journal of Theoretical Probability*, 23(1):1–20, 2010.
- [30] Yong-Qua Yin, Zhi-Dong Bai, and Pathak R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability theory and related fields*, 78(4):509–521, 1988.
- [31] Fang Zhaoben, Liang Ying-Chang, and Bai Zhi-Dong. *Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance Statistics: Random Matrix Theory and Its Applications*. World Scientific, 2014.

# Sažetak

Posljednjih godina, motivirani zahtjevima iz primjena, mnogi matematičari svoja su istraživanja usmjerili na slučajne matrice rastućih dimenzija. U ovom radu opisani su razlozi usmjeravanja pažnje na uzoračke kovarijacijske i korelacijske matrice podataka te pripadne svojstvene vrijednosti i svojstvene vektore.

U prvom poglavlju je napravljen kratak pregled dosadašnjih rezultata na temu asimptotskog ponašanja najvećih svojstvenih vrijednosti slučajnih kovarijacijskih matrica rastućih dimenzija u kontekstu tzv. gaussovskih matrica. Također, napravljen je pregled rezultata u slučaju razdioba s regularno varirajućim repovima, kod kojih je razvoj teorije išao sporije. Naredna poglavlja bave se recentnim rezultatima u slučaju razdioba s regularno varirajućim repovima.

Drugo poglavlje sadrži rezultate najnovijih istraživanja o asimptotskom ponašanju najveće svojstvene vrijednosti uzoračke kovarijacijske matrice za matricu podataka s nezavisno jednako distribuiranim komponentama, čije pripadne distribucije karakterizira težak rep. Također, prikazani su rezultati o svojstvenim vektorima pridruženim najvećim svojstvenim vrijednostima i generalizacija na autokovarijacijske matrice.

U trećem poglavlju uvedena je zavisnost između redaka i stupaca matrice podataka. Nadalje, prikazano je kako se asimptotska konvergencija najveće svojstvene vrijednosti može dobiti iz konvergencije odgovarajućeg točkovnog procesa.

Konačno, u četvrtom poglavlju su ilustrirani rezultati iz prethodnih poglavlja na simuliranim i empirijskim podacima. Na kraju je provedena kratka analiza vremenskih nizova komponenti S&P 500 dioničkog indeksa.

# Summary

In recent years, motivated by the application, many mathematicians have been focusing their research on random matrices of growing dimensions. This Master's Thesis describes the reason why great attention is paid to the sample covariance and correlation data matrices and their eigenvalues and the corresponding eigenvectors.

In the first chapter, in the context of so-called Gaussian data matrices, a brief review of the results on the asymptotic behaviour of the largest eigenvalues and corresponding eigenvectors of random covariance matrices with growing dimensions is presented. Also, an overview of the results in the case of the distributions which satisfy regular variation condition was made, where the development of the theory progressed more slowly. Chapters afterwards contain recent results in the case of distributions which satisfy regular variation condition.

The second chapter contains the results of the newest research on the asymptotic behavior of the largest eigenvalue of the sample covariance matrix for data matrix with independent and identically distributed components, coming from heavy tailed distribution. Also, the results on the eigenvectors associated with the largest eigenvalues, as well as the generalization on the autocovariance matrices, are presented.

The third chapter introduces the dependence between the rows and columns of the data matrix. Furthermore, it is presented that asymptotic convergence of the largest eigenvalue can be obtained from the convergence of the corresponding point processes.

Finally, the illustrations on simulated and empirical data of theoretical results from the preceeding chapters are given in the fourth chapter. Lastly, a brief analysis of the time series components of the S&P 500 stock index was conducted.

# Životopis

Rođen sam 14. travnja 1993. godine u Zagrebu. Nakon završene osnovne škole, 2008. godine, upisujem X. gimnaziju "Ivan Supek". Po završetku srednje škole svoje obrazovanje nastavljam na Prirodoslovno matematičkom fakultetu u Zagrebu, gdje 2012. godine upisujem preddiplomski studij Matematika. 2016. godine, nakon završenog preddiplomskog studija upisujem diplomski studij Financijske i poslovne matematike, također na Prirodoslovno matematičkom fakultetu u Zagrebu. Prvu godinu završavam s odličnim uspjehom, na temelju kojeg u 2017. godini počinjem primati stipendiju za izvrsnost koju dodjeljuje općina Stupnik.